

Decomposition by Successive Convex Approximation: A Unifying Approach for Linear Transceiver Design in Interfering Heterogenous Networks

Mingyi Hong, Qiang Li, Ya-Feng Liu and Zhi-Quan Luo

Abstract

In this work, we study a general downlink linear precoder design problem in a multi-cell heterogenous network (HetNet), in which macro/pico base stations (BSs) are densely deployed within each cell. The problem is formulated in a very general setting as the users' sum-utility maximization problem, in which each user's utility is directly related to its achievable rate. Our formulation includes many practical precoder design problems such as multi-cell coordinated linear precoding, full and partial per-cell coordinated multi-point (ComP) transmission, zero-forcing precoding and joint BS clustering and beamforming/precoding as special cases.

The general sum-utility maximization problem is difficult due to its non-convexity as well as the coupling of all users' precoders through matrix-valued multiuser interference. In this paper, we propose to use a novel convex approximation technique to approximate the original problem by a series convex subproblems, each of which decomposes across all the cells (or BSs). The convexity of the subproblems allows for efficient computation, while their decomposability leads to distributed implementation. Our approach is made possible by the identification of certain key convexity properties of the sum-utility objective. Moreover, in many important network settings, the overall computational complexity can be further reduced by solving, in either an exact or an inexact manner, the per-cell subproblems using customized algorithms. Extensive simulation experiments show that the proposed framework is quite effective for solving interference management problems in many practical settings.

Index Terms

Heterogenous Wireless Networks, Interfering Networks, Convex Decomposition, Successive Convex Approximation, Multicell MIMO Network Resource Allocation

This research is supported by the AFOSR, Grant No. 00008547.

M. Hong and Z.-Q. Luo are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA. Q. Li is with Department of Electronic Engineering, Chinese University of Hong Kong, Hong Kong, China. Y.-F. Liu is with the State Key Lab of Scientific and Engineering Computing, Chinese Academy of Sciences, Beijing, 100190, China.

I. INTRODUCTION

Heterogenous network (HetNet) has recently emerged as a promising wireless network architecture capable of accommodating the explosive demand for wireless data [1]. In HetNet, the upper tier high-power BSs such as Macro BSs provide per-cell interference management as well as blanket coverage, while at the lower tier low-power access points such as micro/pico/femto BSs and relays are densely deployed to provide capacity extension (see Fig. 1). This new paradigm of network design brings the transmitters and receivers closer to each other, thus is able to provide high link quality with low transmission power [2], [3].

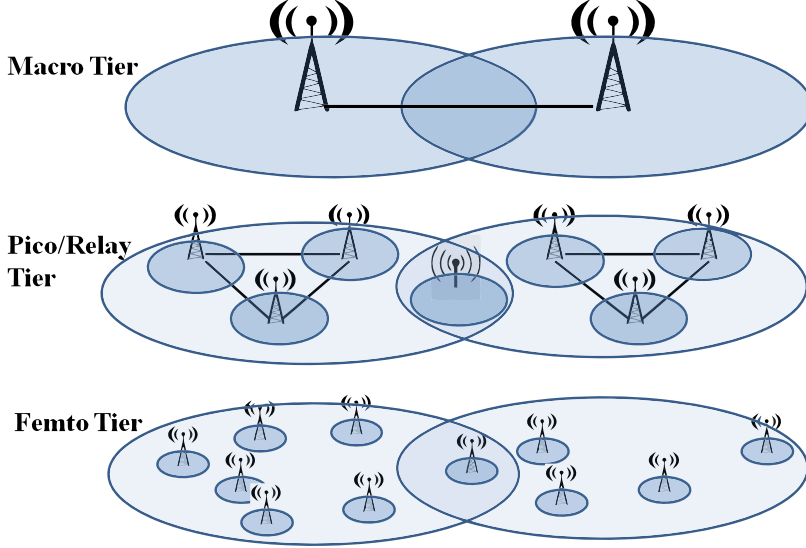


Fig. 1. The Multi-Tiered Structure of the HetNet.

Due to the large number of potential interfering nodes in the network, one of the key challenges in the design of effective resource allocation schemes in the HetNet is to properly mitigate both the inter-cell and intra-cell multiuser interference. Interference management for the HetNet, or for general interference networks, has been under extensive research recently [4], [5]. Introducing appropriate coordination among the nodes in the network, either in the physical layer or in higher layers, has been shown as an effective means for such purpose [6]–[8]. For the network with the nodes equipped with multiple antennas, major approaches for physical layer coordinated interference management include coordinated beamforming (CB) and joint processing (JP) [7]. The CB approach allows the nodes to coordinate in the beamformer/precoder level. By a joint design of the transmit/receive beamformers/precoders of all the nodes in the network, excessive interference can be avoided [9]–[14]. The JP approach, also known as the coordinated multi-point (ComP) transmission, optimizes the transceiver structures assuming that the users' data are available at all the BSs. For example, in a downlink network, a single virtual BS can be formed if the users' data is shared among all transmitting BSs. In this way, transmission scheme designed for single cell MIMO broadcasting channels, such as non-linear dirty-paper coding (DPC) [15]–[18] and linear precoding schemes [19]–[22], can be used for coordinated transmission. However, due to the high signaling overhead associated with implementing full JP for all the BSs in the network, in the HetNet setting, a combination of these two approaches is usually adopted.

For example, JP can be performed in a per-cell basis to cancel intra-cell interference, while the CB is used to mitigate inter-cell interference [6], [20].

Regardless of the coordination schemes used, interference management is usually formulated into problems that optimize certain system utility functions, which are directly related to the users' individual rates [4]. These utility functions, when chosen properly, can well balance the network spectrum efficiency and the fairness level among the users. Unfortunately, it has been recently established that, for a large class of utility functions, the associated optimization problems are difficult to solve (except for a few special cases, see [14], [23]–[25]). As a result, many practical low-complexity algorithms that compute high-quality solutions to the interference management problems have been recently developed. The key to designing practical algorithms for interference management is to recognize certain convexity and/or decomposability of the underlying utility maximization problems. Convexity leads to efficient computational algorithms, while decomposability is crucial for distributed implementation, especially in large-scale networks [26], [27].

Reference [28] is the first to recognize that in a MIMO interfering channel (IC), individual users' achievable rate is *concave* in its own transmit covariance while at the same time *convex* in all other interfering users' transmit covariances. Similar concave-convex properties have later been established in other interference networks such as the multi-cell OFDMA networks [29], and has since been leveraged heavily to design resource allocation algorithms in various network settings [9], [11], [25], [29]–[34]. The convex-concave property allows one to obtain, for each individual user, an approximated version of the sum-rate objective by linearizing its convex part while keeping its concave part unchanged. In this way, the users can successively optimize their transmit strategies by solving a series of convex subproblems. Different convex approximation approaches that are not based on the convex-concave properties are also possible, see e.g., [14], [35]. However, all the schemes cited above do not decompose well across the nodes: for the schemes based on the convex-concave structure, the convexification procedures can only be done for one node at each time, thus only *a single* node can update its transmit strategy in each iteration; for the algorithm proposed in [35], the convex subproblem is still coupled among all the users. Moreover, they are mainly designed for peer-to-peer networks, with each transmitter dedicated to transmit to a single receiver. Hence they are not suitable for the HetNet setting where each BS can transmit to multiple users while at the same time each users can receive from multiple BSs as well. There are a few recent works that have attempted to address these drawbacks [36]–[39]. However, there is no theoretical convergence analysis for these algorithms.

Nevertheless, decomposability structure of interference management problems is highly desirable. When judiciously exploited, it can lead to efficient distributed implementation. This fact has long been recognized for other important large-scale network optimization problems such as the network utility maximization (NUM) problems (which are a class of problems with convex, separable objectives and coupled constraints). See [27] for a recent survey on various techniques to achieve decomposability across all the users for the NUM-related problems. However, unlike the NUM problems, in the interfering networks the nodes are tightly coupled in a nonlinear manner through multi-user interference. As a result, even in relatively simple network setting with a set of single

antenna transceiver pairs communicating simultaneously (i.e., the interference channel model), decomposability structure is difficult to come by. In this case, a simple way to achieve decomposability is to allow each transceiver pairs to optimize completely by their own, while disregarding the interference they generate to other nodes in the system [40]–[43]. Despite their simplicity, this type of algorithms suffer from non-convergence and low throughput when the multi-user interference becomes high. Recently, a weighted minimum mean square error (WMMSE) algorithm is proposed in [44] for general utility optimization problems in interfering broadcast channels (IBC). Surprisingly, in each of its steps, computation is completely decoupled among the interfering BSs in network. Such decomposition is achieved by establishing a key equivalence relationship between the utility maximization problem and a weighted MSE minimization problem. Local optimal solutions of the latter problem can be obtained via solving three subproblems alternatingly, each of which can be decoupled completely across the users. However, it is not clear if the equivalence relationship derived still holds true for general HetNet setting.

To the best of our knowledge, to this point there is no general approach for decomposing interference-coupled utility maximization problems in HetNet, or in general interference networks. In this work, we propose to achieve decomposability by means of successive convex approximation. Central to our approach is a key observation that reveals certain hidden convexity for a wide range of sum-utility maximization problems. The identified property allows us to approximate the original non-convex problem by a series of convex subproblems, each of which is completely decoupled among the nodes in the network. Based on different ways in which the convex subproblems are solved, two low-complexity algorithms are proposed, each bearing wide applicability in interference management problems.

The rest of the paper is organized as follows. In Section II, we outline a general system model for interference management in HetNet, and describe its applicability in many important practical problems. In Section III, we present a key convexity structure of the considered utility maximization problem, which leads to a general successive convex approximation algorithm. In Section IV, the proposed general algorithm is specialized to different interference management scenarios. In Section V, a useful extension of the algorithm is developed, and its application is discussed. Numerical results are given in Section VI, and concluding remarks are provided in Section VII.

Notations: For a symmetric matrix \mathbf{X} , $\mathbf{X} \succeq 0$ signifies that \mathbf{X} is positive semi-definite. We use $\text{Tr}[\mathbf{X}]$, $|\mathbf{X}|$, \mathbf{X}^H , $\rho(\mathbf{X})$ and $\text{Rank}(\mathbf{X})$ to denote the trace, determinant, Hermitian, spectral radius and the rank of a matrix, respectively. The (m, n) -th element of a matrix \mathbf{X} is denoted by $\mathbf{X}[m, n]$. We use \mathbf{I}_n to denote a $n \times n$ identity matrix. Moreover, we let $\mathbb{R}^{N \times M}$ and $\mathbb{C}^{N \times M}$ denote the set of real and complex $N \times M$ matrices, and use \mathbb{S}^N , \mathbb{S}_+^N , \mathbb{S}_{++}^N to denote the set of $N \times N$ Hermitian, Hermitian positive semi-definite and Hermitian positive definite matrices, respectively. Finally, we use the notation $0 \leq a \perp b \leq 0$ to indicate $a \geq 0, b \geq 0, a \times b = 0$. The main notations used in this paper are listed in Table I.

TABLE I
A LIST OF NOTATIONS

Symbols	Description
\mathcal{K}	The set of cells
\mathcal{I}	The set of all users
\mathcal{Q}	The set of all BSs
\mathcal{Q}_k	The set of BSs in cell k
\mathcal{I}_k	The set of users in cell k
M	The number of transmit antennas per BS
N	The number of receive antennas per user
d_{i_k}	The number of transmitted data streams for user i_k
$\mathbf{H}_{i_k}^{q_\ell}$	The channel between user i_k and BS q_ℓ
$\mathbf{V}_{i_k}^{q_k}$	The transmit beamformer from BS q_k to user i_k
\mathbf{E}_{i_k}	The receive MSE matrix for user i_k

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a downlink multi-cell HetNet consists of a set $\mathcal{K} \triangleq \{1, \dots, K\}$ of cells. Within each cell k there is a set of $\mathcal{Q}_k = \{1, \dots, Q_k\}$ distributed base stations (BS) such as macro/micro/pico BSs which provide service to users located in different areas of the cell. Assume that in each cell k , there is low-latency backhaul network connecting the set of BSs \mathcal{Q}_k to a central controller (usually the macro BS), who makes the resource allocation decisions for all BSs within the cell. Furthermore, this central entity has access to the data signals of all the users in its cell. Let $\mathcal{I}_k \triangleq \{1, \dots, I_k\}$ denote the users associated with cell k . Each of the users $i_k \in \mathcal{I}_k$ is served jointly by a subset of BSs in \mathcal{Q}_k . Let \mathcal{I} denote the set of all the users and let \mathcal{Q} denote the set of all the BSs, respectively. Assume that each BS has M transmit antennas, and each user has N receive antennas. Let $\mathbf{H}_{i_k}^{q_\ell} \in \mathbb{C}^{N \times M}$ denote the channel matrix between the q -th BS in the ℓ -th cell and the i -th user in the k -th cell. Similarly, we use $\mathbf{H}_{i_k}^\ell$ to denote the channel matrix between all the BSs in the ℓ -th cell to the user i_k , i.e., $\mathbf{H}_{i_k}^\ell \triangleq \{\mathbf{H}_{i_k}^{q_\ell}\}_{q_\ell \in \mathcal{Q}_\ell} \in \mathbb{C}^{N \times M Q_\ell}$.

Suppose that it is possible to transmit $d_{i_k} \leq \min\{M, N\}$ parallel data streams to user i_k . Let $\mathbf{V}_{i_k}^{q_k} \in \mathbb{C}^{M \times d_{i_k}}$ denote the transmit precoder that BS q_k uses to transmit data $\mathbf{s}_{i_k} \in \mathbb{C}^{d_{i_k}}$ to user i_k . Define $\mathbf{V}_{i_k} \triangleq \{\mathbf{V}_{i_k}^{q_k}\}_{q_k \in \mathcal{Q}_k} \in \mathbb{C}^{M Q_k \times d_{i_k}}$, $\mathbf{V}^{q_k} \triangleq \{\mathbf{V}_{i_k}^{q_k}\}_{i_k \in \mathcal{I}_k} \in \mathbb{C}^{M I_k \times d_{i_k}}$ and $\mathbf{V}^k \triangleq \{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}_k}$ as the collection of all precoders intended for user i_k , the collection of all the precoder belong to BS q_k , and the collection of all precoders in cell k , respectively. Let $\mathbf{V} \triangleq \{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}}$.

Using the above definitions, we can express the transmitted signal of BS q_k as well as the combined transmitted signal for all the BSs in cell k as:

$$\mathbf{x}^{q_k} = \sum_{i_k \in \mathcal{I}_k} \mathbf{V}_{i_k}^{q_k} \mathbf{s}_{i_k} \in \mathbb{C}^{M \times 1}, \quad \mathbf{x}^k = \sum_{i_k \in \mathcal{I}_k} \mathbf{V}_{i_k} \mathbf{s}_{i_k} \in \mathbb{C}^{M Q_k \times 1}.$$

The received signal $\mathbf{y}_{i_k} \in \mathbb{C}^{N \times 1}$ of user i_k is

$$\mathbf{y}_{i_k} = \sum_{\ell \in \mathcal{K}} \mathbf{H}_{i_k}^\ell \mathbf{x}^\ell + \mathbf{z}_{i_k} = \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \mathbf{s}_{i_k} + \underbrace{\sum_{j_k \neq i_k} \mathbf{H}_{i_k}^k \mathbf{V}_{j_k} \mathbf{s}_{j_k}}_{\text{intra-cell interference}} + \underbrace{\sum_{\ell \neq k} \sum_{j_\ell \in \mathcal{I}_\ell} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j_\ell} \mathbf{s}_{j_\ell}}_{\text{inter-cell interference}} + \mathbf{z}_{i_k} \quad (1)$$

where $\mathbf{z}_{i_k} \in \mathbb{C}^{N \times 1}$ is the additive white complex Gaussian noise with distribution $\mathcal{CN}(0, \sigma_{i_k}^2 \mathbf{I}_N)$.

Let $\mathbf{U}_{i_k} \in \mathbb{C}^{N \times d_{i_k}}$ denote the linear receiver used by user i_k to decode the intended signal. Then the estimated signal for user i_k is: $\hat{\mathbf{s}}_{i_k} = \mathbf{U}_{i_k}^H \mathbf{y}_{i_k}$. The mean square error (MSE) for user i_k can be written as

$$\begin{aligned} \mathbf{E}_{i_k} &\triangleq \mathbb{E}[(\mathbf{s}_{i_k} - \hat{\mathbf{s}}_{i_k})(\mathbf{s}_{i_k} - \hat{\mathbf{s}}_{i_k})^H] \\ &= (\mathbf{I}_{d_{i_k}} - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k})(\mathbf{I}_{d_{i_k}} - \mathbf{U}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k})^H + \sum_{(\ell, j) \neq (k, i)} \mathbf{U}_{i_k}^H \mathbf{H}_{i_k}^\ell \mathbf{V}_{j\ell} \mathbf{V}_{j\ell}^H (\mathbf{H}_{i_k}^\ell)^H \mathbf{U}_{i_k} + \sigma_{i_k}^2 \mathbf{U}_{i_k}^H \mathbf{U}_{i_k}. \end{aligned} \quad (2)$$

The Minimum MSE (MMSE) receiver minimizes user i_k 's MSE, and can be expressed as [45]

$$\mathbf{U}_{i_k}^{\text{mmse}} = \left(\sum_{(\ell, j)} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j\ell} (\mathbf{V}_{j\ell})^H (\mathbf{H}_{i_k}^\ell)^H + \sigma_{i_k}^2 \mathbf{I}_N \right)^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \triangleq \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \quad (3)$$

where $\mathbf{C}_{i_k} \in \mathbb{S}_{++}^N$ denotes user i_k 's received signal covariance matrix. When the MMSE receiver is used, the MMSE matrix (2) is reduced to

$$\mathbf{E}_{i_k}^{\text{mmse}} = \mathbf{I}_{d_{i_k}} - \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \succeq 0. \quad (4)$$

Clearly we also have $\mathbf{I}_{d_{i_k}} - \mathbf{E}_{i_k}^{\text{mmse}} \succeq 0$.

Let us assume that Gaussian signaling is used and the interference is treated as noise. Further assume that all the BSs in cell k form a single virtual BS that jointly transmit to user $i_k \in \mathcal{I}_k$, then \mathbf{V}_{i_k} can be viewed as the virtual precoder for user i_k . The achievable rate for user i_k is given by [46]

$$R_{i_k} = \log \left| \mathbf{I}_N + \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \left(\sum_{(\ell, j) \neq (k, i)} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j\ell} \mathbf{V}_{j\ell}^H (\mathbf{H}_{i_k}^\ell)^H + \sigma_{i_k}^2 \mathbf{I}_N \right)^{-1} \right| \quad (5)$$

$$= -\log |\mathbf{E}_{i_k}^{\text{mmse}}| \quad (6)$$

where the last equality is the well-known relationship between the transmission rate and the MMSE matrix (see e.g., [22]). It can be derived using the matrix inversion lemma. We will occasionally use the notations $R_{i_k}(\mathbf{V})$, $\mathbf{C}_{i_k}(\mathbf{V})$ and $\mathbf{E}_{i_k}^{\text{mmse}}(\mathbf{V})$ to make their dependencies on \mathbf{V} explicit.

Let $f_{i_k}(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}$ be the utility function of user i_k 's data rate. We make the following assumptions on the function $f_{i_k}(\cdot)$:

- A-1)** $f_{i_k}(x)$ is a concave non-decreasing function in x for all $x \geq 0$;
- A-2)** $f_{i_k}(-\log(|\mathbf{X}|))$ is convex in \mathbf{X} , For all $\mathbf{I} \succeq \mathbf{X} \succeq 0$;
- A-3)** $f_{i_k}(x)$ is continuously differentiable (i.e., a smooth function).

Note that this family of utility functions includes well known utilities such as the weighted sum rate, the geometric mean of one plus rate and the harmonic mean rate utility functions (see [44]). They differ considerably with those studied in references [47]–[51] which, although admit concave representations, are not directly related to individual users' rates.

Let $s_{i_k}^{q_k}(\cdot)$ be a penalty term of the precoder $\mathbf{V}_{i_k}^{q_k}$. We make the following assumptions on the function $s_{i_k}^{q_k}(\cdot)$:

B-1) $s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k})$ is a convex function.

B-2) $s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k})$ is a function that is continuous, but possibly nonsmooth.

In this paper, we consider the general system-level sum utilities maximization problem in the following form

$$\begin{aligned}
 \max \quad & u(\mathbf{V}) \\
 \text{s.t.} \quad & u(\mathbf{V}) = f(\mathbf{V}) - s(\mathbf{V}) \\
 & f(\mathbf{V}) = \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} f_{i_k}(R_{i_k}(\mathbf{V})) \\
 & s(\mathbf{V}) = \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \sum_{q_k \in \mathcal{Q}_k} s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k}) \\
 & \mathbf{V}^{q_k} \in \mathcal{V}^{q_k} \quad \forall q_k \in \mathcal{Q}, \quad \mathbf{V}^k \in \mathcal{V}^k \quad \forall k \in \mathcal{K}
 \end{aligned} \tag{SYSTEM}$$

where \mathcal{V}^{q_k} and \mathcal{V}^k represent the feasible sets for BS q_k 's transmit precoder and the collection of all precoders in cell k , respectively. Let \mathcal{V} denote the feasible set for \mathbf{V} . The penalty term in the objective, when properly designed, can induce certain desired structure in the resulting precoding matrices.

When the functions $f(\mathbf{V})$ and $s(\mathbf{V})$ as well as the feasible sets $\{\mathcal{V}^{q_k}\}$ and $\{\mathcal{V}^k\}$ are properly specified, the general problem (SYSTEM) can cover a wide range of transceiver design problems in multicell networks. In the following, we list several of those problems that are of wide interests.

- **MIMO IBC/IMAC/IC channels with inter-BS CB** [9], [10], [14], [22], [39], [44], [52]: Each cell k has a single BS serving all the users \mathcal{I}_k . In this case no penalty term $s(\mathbf{V})$ is assumed in the objective, and the constraint set \mathcal{V}^{q_k} becomes the same as the constraint set \mathcal{V}^k , which is given by the following sum-power constrained set

$$\mathcal{V}^k = \left\{ \mathbf{V}^k : \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H] \leq \bar{P}_k \right\}. \tag{7}$$

- **Multicell MIMO network with intra-cell ComP and inter-cell CB** [53], [54]: The BSs in different cells cooperate in the precoder level, while the BSs in the same cell share the user data and perform joint transmission. Each BS q_k has a separate power constraint:

$$\mathcal{V}^{q_k} = \left\{ \mathbf{V}^{q_k} : \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^{q_k} (\mathbf{V}_{i_k}^{q_k})^H] \leq \bar{P}^{q_k} \right\}. \tag{8}$$

This model generalizes those for the linear precoder design in the IBC model (e.g., [39], [44]), as when we view all the BSs Q_k in a cell k as a giant virtual BS, we recover the precoder design problem in IBC model, except that the sum-power constraint (7) is replaced by the set of *per group of antenna* power constraints (each small BS q_k is viewed as a group of antennas of the virtual transmitter k).

- **Multicell MIMO network with intra-cell partial ComP and inter-cell CB** [54]–[56]: As is well known from the literature (see e.g., [6], [7]), performing full ComP in each cell can achieve huge improvement of the

overall spectrum efficiency, while suffering from high level of overhead in the backhaul network. A practical alternative scheme to tradeoff spectrum efficiency with acceptable overhead is to implement a *partial Comp* strategy, in which each user is served by not all, but only a few of BSs in each cell. In this case, the BSs in different cells cooperate on the precoder level, while the BSs in the same cell are grouped into different (possibly overlapping) clusters with small sizes, within which they fully cooperate for transmission (see Fig. 2 for an illustration). In this case, besides precoder design, the cluster membership of the BSs needs to be decided. This task can be done *jointly* with precoder design by properly specifying the penalty term $s(\mathbf{V})$. The requirement that each user is served by a few BSs translates to the restriction that for each BS $q_k \in \mathcal{Q}_k$, its precoder $\mathbf{V}^{q_k} = \{\mathbf{V}_{i_k}^{q_k}\}_{i_k \in \mathcal{I}_k}$ should contain only a few nonzero block components [54]. To induce such block sparsity, the penalty term can take the form of

$$s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k}) = \gamma_{i_k}^{q_k} \|\mathbf{V}_{i_k}^{q_k}\|_2, \quad (9)$$

with $\gamma_{i_k}^{q_k} \geq 0$ being some constant. See [54] and the reference therein for motivation of using (9). The constraint set for this problem is the same as that in (8).

- **Single cell MIMO network with intra-cell ZF** [21]: All the BSs in the cell (say cell 1) jointly perform ZF precoding. Such precoding technique is referred to as block-diagonalization (BD) precoding [19]. That is, the designed precoders ensure that each user receives the intended transmission free of interference. The penalty term $s(\mathbf{V})$ is assumed to be not present. The feasible sets become

$$\mathcal{V}^1 = \left\{ \mathbf{V}^1 : \mathbf{H}_{j_1}^1 \mathbf{V}_{i_1} (\mathbf{V}_{i_1})^H (\mathbf{H}_{j_1}^1)^H = \mathbf{0}, \forall j_1 \neq i_1, j_1, i_1 \in \mathcal{I}_1, \right\} \quad (10)$$

$$\mathcal{V}^{q_1} = \left\{ \mathbf{V}^{q_1} : \sum_{i_1 \in \mathcal{I}_1} \text{Tr}[\mathbf{V}_{i_1}^{q_1} (\mathbf{V}_{i_1}^{q_1})^H] \leq \bar{P}^{q_1}, \forall q_1 \in \mathcal{Q}_1 \right\}. \quad (11)$$

We should note that the ZF scheme is a special case of the linear precoding schemes discussed above, as it imposes the additional “zero-forcing” structure on the precoders to be designed.

- **Multicell MIMO network with intra-cell ZF and inter-cell CB** [20]: In this setting, all the BSs in the same cell jointly perform BD-ZF, while the BSs in different cells perform CB. The feasible sets are given as

$$\mathcal{V}^k = \left\{ \mathbf{V}^k : \mathbf{H}_{j_k}^k \mathbf{V}_{i_k} (\mathbf{V}_{i_k})^H (\mathbf{H}_{j_k}^k)^H = \mathbf{0}, \forall j_k \neq i_k, j_k, i_k \in \mathcal{I}_k, \right\} \quad (12)$$

$$\mathcal{V}^{q_k} = \left\{ \mathbf{V}^{q_k} : \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^{q_k} (\mathbf{V}_{i_k}^{q_k})^H] \leq \bar{P}^{q_k}, \forall q_k \in \mathcal{Q}_k \right\}. \quad (13)$$

- **Interfering OFDMA network** [29], [31], [37], [41]: Assuming that each cell k has only a single BS that serves a single user (denoted as user k). Suppose there is a total of N orthogonal channels that can be used by all the cells. Then the traditional power control problem in interfering OFDMA network is a special case of the precoder design problem by setting \mathbf{V}_k and \mathbf{H}_k^ℓ to be diagonal for all $\ell, k \in \mathcal{K}$, and let $M = d_k = N$

for all k . In this way, user k 's rate becomes

$$R_k = \sum_{n=1}^N \log \left(1 + \frac{|\mathbf{H}_k^k[n, n]|^2 |\mathbf{V}_k[n, n]|^2}{\sigma_k^2 + \sum_{\ell \neq k} |\mathbf{H}_k^\ell[n, n]|^2 |\mathbf{V}_\ell[n, n]|^2} \right) \quad (14)$$

where $|\mathbf{H}_k^\ell[n, n]|^2$ represents the channel gain on channel n from BS ℓ to user k ; $|\mathbf{V}_k[n, n]|^2$ represents the transmit power for BS k on channel n .

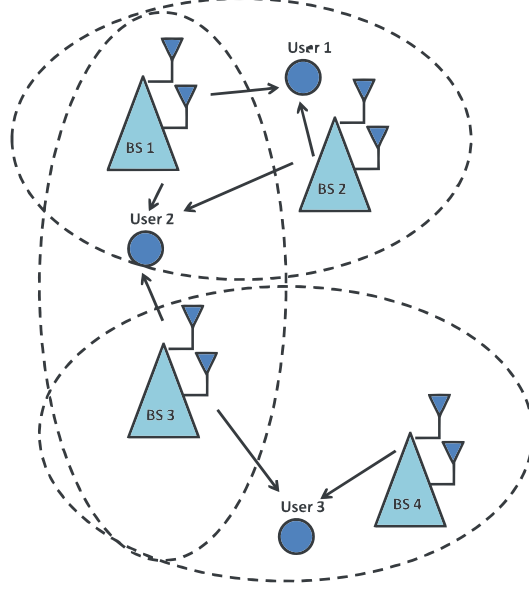


Fig. 2. A graphical illustration for the partial ComP with intra-cell BS clustering. In this figure, three overlapping groups are formed, which respectively contains BSs (1, 2), (1, 3) and (3, 4).

Despite the wide applicability of (SYSTEM), solving it to its global optimality is often very difficult, even without the penalty term $s(\mathbf{V})$. In fact, a set of recent works [14], [23], [57], [58] have rigorously established the level of difficulties of various forms of problem (SYSTEM) using the computational complexity theory [59]. The main message is that except for a very few cases such as MISO min utility maximization problem (with $N = 1$ for all the users, which can be equivalently transformed to a convex problem), solving the problem (SYSTEM) is generally NP-hard. We refer the readers to [4] for a summary of these complexity results. The NP-hardness of the problem indicates that it is not even possible to find an equivalent convex reformulation for it. Thus the best that one can do is to seek efficient low-complexity algorithms that provide approximately optimal solutions. Unfortunately, due to the fact that the variables $\{\mathbf{V}^k\}_{k=1}^K$ that belong to different cells are tightly coupled in the objective function via multi-user interference, even finding efficient (preferably distributed) approximate algorithm for this family of problems can be challenging.

III. A GENERAL SUCCESSIVE CONVEX APPROXIMATION APPROACH

In this section, we will present our main approach for computing a high quality solution for the general problem (SYSTEM). As stated in the previous section, solving problem (SYSTEM) directly is difficult due to its non-convexity. Our approach is to instead solve a series of convex subproblems, each of which is a local approximation

of (SYSTEM). A desirable feature of our approach is that each of the obtained subproblems is not only convex with respect to the transmit precoder \mathbf{V} , but is completely *separable* among the precoders of different cells. Consequently, the computation of the resource allocation in different stages of the algorithm can be carried out independently in a parallel fashion by individual cells.

A. A Local Approximation for (SYSTEM)

To approximate the objective function of (SYSTEM), we begin by deriving a simple local approximation for the individual users' utility function $f_{i_k}(\cdot)$ using the convexity assumed in assumption A-2. More specifically, let $\hat{\mathbf{V}} \in \mathcal{V}$ be a feasible solution to problem (SYSTEM). Let $\hat{\mathbf{E}}_{i_k} \triangleq \mathbf{E}_{i_k}^{\text{mmse}}(\hat{\mathbf{V}})$ denote the MMSE evaluated at $\hat{\mathbf{V}}$. From the relationship of the individual users' rates and their MMSE matrices (6), we can express $f_{i_k}(\cdot)$ as a function of $\mathbf{E}_{i_k}^{\text{mmse}}$ only.

Using Assumption A-2), we have

$$\begin{aligned} f_{i_k}(\mathbf{V}) &= f_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}) = f_{i_k}(-\log |\mathbf{E}_{i_k}^{\text{mmse}}|) \\ &\geq f_{i_k}(-\log |\hat{\mathbf{E}}_{i_k}|) - \frac{\partial f_{i_k}(x)}{\partial x} \Big|_{x=-\log |\hat{\mathbf{E}}_{i_k}|} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{E}_{i_k}^{\text{mmse}} - \hat{\mathbf{E}}_{i_k}) \right] \\ &\triangleq \hat{a}_{i_k} - \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{E}_{i_k}^{\text{mmse}} - \hat{\mathbf{E}}_{i_k}) \right] \\ &\triangleq \bar{h}_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}; \hat{\mathbf{E}}_{i_k}) \end{aligned} \quad (15)$$

where the inequality is due to the property of the convex function; \hat{c}_{i_k} and \hat{a}_{i_k} are two constants that are not related to $\mathbf{E}_{i_k}^{\text{mmse}}$ or \mathbf{V} , and $\hat{c}_{i_k} \geq 0$ due to the non-decreasing property of $f(\cdot)$ assumed in A-1).

Note that the function $\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}; \hat{\mathbf{E}}_{i_k})$ is a locally approximated version of $f_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}})$ at the point $\hat{\mathbf{E}}_{i_k}$. In fact, the approximation is a global underestimate of $f_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}})$, and it is tight at the point $\hat{\mathbf{E}}_{i_k}$. That is, $\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}; \hat{\mathbf{E}}_{i_k}) \leq f_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}})$ for all feasible $\mathbf{E}_{i_k}^{\text{mmse}}$ and $\hat{\mathbf{E}}_{i_k}$, and that $\bar{h}_{i_k}(\hat{\mathbf{E}}_{i_k}; \hat{\mathbf{E}}_{i_k}) = f_{i_k}(\hat{\mathbf{E}}_{i_k})$.

Unfortunately, the above approximation does not seem to simplify the problem. Although the sum of the derived approximated functions, $\sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \bar{h}_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}; \hat{\mathbf{E}}_{i_k})$, can be viewed as a locally approximated version of $f(\mathbf{V})$, and it is indeed convex (in fact linear) with respect to $\{\mathbf{E}_{i_k}^{\text{mmse}}\}$, it is still non-convex with respect to the system precoder \mathbf{V} (cf. (4)). Our next step is therefore to further approximate $\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}; \hat{\mathbf{E}}_{i_k})$ by a convex function of \mathbf{V} . To this end, we need a key technical lemma that explores some hidden convexity property of the function $\bar{h}_{i_k}(\cdot)$.

Lemma 1 *The function $\text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \right]$ is jointly convex with respect to the pair of variables $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k})$ in the feasible region $(\mathbb{C}^{MQ_k \times d_{i_k}}, \mathbb{S}_{++}^N)$.*

Proof: Let us define

$$l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}) \triangleq \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \right].$$

Consider the epigraph of $l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k})$, i.e.,

$$\{(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t) \mid l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}) \leq t\} = \left\{(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t) \mid \text{Tr} \left[\hat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \hat{\mathbf{E}}_{i_k}^{-1/2} \right] \leq t \right\}.$$

It suffices to show that the epigraph $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t)$ is a convex set [60, Chapter 3]. To this end, let us consider the following extended set (with $\mathbf{Z}_{i_k} \succeq \mathbf{0}$ being a slack variable):

$$\left\{(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, \mathbf{Z}_{i_k}, t) \mid \text{Tr}[\mathbf{Z}_{i_k}] \leq t, \mathbf{Z}_{i_k} \succeq \hat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \hat{\mathbf{E}}_{i_k}^{-1/2}, \mathbf{Z}_{i_k} \succeq \mathbf{0}\right\}. \quad (16)$$

It is not hard to show that $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t)$ is just a projection of the set defined by (16). Therefore, if the extended set $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, \mathbf{Z}_{i_k}, t)$ is convex, then $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, t)$ is also convex. By applying Schur's complement to $\mathbf{Z}_{i_k} \succeq \hat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \hat{\mathbf{E}}_{i_k}^{-1/2}$, we have

$$\mathbf{Z}_{i_k} \succeq \hat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \hat{\mathbf{E}}_{i_k}^{-1/2} \iff \begin{bmatrix} \mathbf{Z}_{i_k} & \hat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \\ \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \hat{\mathbf{E}}_{i_k}^{-1/2} & \mathbf{C}_{i_k} \end{bmatrix} \succeq \mathbf{0}. \quad (17)$$

Substituting (17) into (16) yields

$$\left\{(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}, \mathbf{Z}_{i_k}, t) \mid \text{Tr}[\mathbf{Z}_{i_k}] \leq t, \mathbf{Z}_{i_k} \succeq \mathbf{0}, \begin{bmatrix} \mathbf{Z}_{i_k} & \hat{\mathbf{E}}_{i_k}^{-1/2} \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \\ \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \hat{\mathbf{E}}_{i_k}^{-1/2} & \mathbf{C}_{i_k} \end{bmatrix} \succeq \mathbf{0}\right\},$$

which is apparently a convex set whenever $\mathbf{C}_{i_k} \succeq \mathbf{0}$.

We remark that a direct proof of this lemma can be obtained by evaluating the second order directional derivative of the function $l_{i_k}(\cdot)$. However this direct approach is not as concise, and we include the proof in the Appendix A for completeness. ■

The result provided by Lemma 1 allows us to further approximate the function $\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}, \hat{\mathbf{E}}_{i_k})$. Specifically, define $\hat{\mathbf{C}}_{i_k} = \mathbf{C}_{i_k}(\hat{\mathbf{V}})$ as user i_k 's received covariance evaluated at point $\hat{\mathbf{V}}$, then we have

$$\begin{aligned} l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k}) &= \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \right] \\ &\geq \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\hat{\mathbf{V}}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k \hat{\mathbf{V}}_{i_k} \right] + \left. \frac{dl_{i_k}(\hat{\mathbf{V}}_{i_k} + t(\mathbf{V}_{i_k} - \hat{\mathbf{V}}_{i_k}), \hat{\mathbf{C}}_{i_k})}{dt} \right|_{t=0} + \left. \frac{dl_{i_k}(\hat{\mathbf{V}}_{i_k}, \hat{\mathbf{C}}_{i_k} + t(\mathbf{C}_{i_k} - \hat{\mathbf{C}}_{i_k}))}{dt} \right|_{t=0} \\ &= \text{Tr}[(\hat{\mathbf{E}}_{i_k})^{-1}] - d_{i_k} + \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} \hat{\mathbf{V}}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} - \hat{\mathbf{V}}_{i_k}) \right] + \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} - \hat{\mathbf{V}}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k \hat{\mathbf{V}}_{i_k} \right] \\ &\quad - \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} \hat{\mathbf{V}}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{C}}_{i_k}^{-1} (\mathbf{C}_{i_k} - \hat{\mathbf{C}}_{i_k}) \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k \hat{\mathbf{V}}_{i_k} \right] \end{aligned}$$

where the inequality is due to the property of the convex function, and the directional derivatives can be obtained similarly as those given in Appendix A. The above inequality combined with the definition of the MMSE matrix

in (4) yields:

$$\begin{aligned}
\bar{h}_{i_k}(\mathbf{E}_{i_k}^{\text{mmse}}; \hat{\mathbf{E}}_{i_k}) &= \hat{a}_{i_k} - \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{E}_{i_k}^{\text{mmse}} - \hat{\mathbf{E}}_{i_k}) \right] \\
&\stackrel{(i)}{=} \hat{a}_{i_k} + \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \mathbf{C}_{i_k}^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} \right] + \hat{c}_{i_k} d_{i_k} - \hat{c}_{i_k} \text{Tr}[\hat{\mathbf{E}}_{i_k}^{-1}] \\
&\geq \hat{a}_{i_k} + \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} \hat{\mathbf{V}}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} - \hat{\mathbf{V}}_{i_k}) \right] + \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} - \hat{\mathbf{V}}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k \hat{\mathbf{V}}_{i_k} \right] \\
&\quad - \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} \hat{\mathbf{V}}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{C}}_{i_k}^{-1} (\mathbf{C}_{i_k} - \hat{\mathbf{C}}_{i_k}) \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k \hat{\mathbf{V}}_{i_k} \right] \\
&= \tilde{a}_{i_k} + \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} \left(\hat{\mathbf{U}}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} + \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} - \hat{\mathbf{U}}_{i_k}^H \left(\sum_{(\ell,j)} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j_\ell} (\mathbf{V}_{j_\ell})^H (\mathbf{H}_{i_k}^\ell)^H \right) \hat{\mathbf{U}}_{i_k} \right) \right] \\
&\triangleq h_{i_k}(\mathbf{V}; \hat{\mathbf{V}})
\end{aligned} \tag{18}$$

where in (i) we have used the definition of the MMSE matrix (4); in the last equality, we have defined $\hat{\mathbf{U}}_{i_k} \triangleq \hat{\mathbf{C}}_{i_k}^{-1} \mathbf{H}_{i_k}^k \hat{\mathbf{V}}_{i_k}$, and \tilde{a}_{i_k} is a constant that is not dependent on \mathbf{V} . Interestingly, $\hat{\mathbf{U}}_{i_k}$ defined in this way is in fact the MMSE receiver for user i_k when the system precoder is given by $\hat{\mathbf{V}}$ (cf. (3)).

Combining (15) and (18), we see that $h_{i_k}(\mathbf{V}; \hat{\mathbf{V}})$ is in fact a locally approximated version of $f_{i_k}(\mathbf{V})$ that satisfies the following two properties for all feasible \mathbf{V} and $\hat{\mathbf{V}}$:

$$h_{i_k}(\hat{\mathbf{V}}; \hat{\mathbf{V}}) = f_{i_k}(\hat{\mathbf{V}}), \quad h_{i_k}(\mathbf{V}; \hat{\mathbf{V}}) \leq f_{i_k}(\mathbf{V}). \tag{19}$$

Let us define $h(\mathbf{V}; \hat{\mathbf{V}}) \triangleq \sum_{i_k \in \mathcal{I}_k} h_{i_k}(\mathbf{V}; \hat{\mathbf{V}})$. The above results further imply that for all feasible \mathbf{V} , $\hat{\mathbf{V}}$,

$$h(\hat{\mathbf{V}}; \hat{\mathbf{V}}) = f(\hat{\mathbf{V}}) \tag{20a}$$

$$h(\mathbf{V}; \hat{\mathbf{V}}) \leq f(\mathbf{V}). \tag{20b}$$

That is, the function $h(\mathbf{V}; \hat{\mathbf{V}})$ is a locally tight, *universal* lower bound for the sum-utility function $f(\mathbf{V})$. A direct consequence of this observation is that the function $h(\mathbf{V}; \hat{\mathbf{V}}) - s(\mathbf{V})$ is a universal lower bound for $u(\mathbf{V})$. Importantly, $h(\mathbf{V}; \hat{\mathbf{V}})$ is in fact a *concave* function with respect to the variables \mathbf{V} . To see this, we can expand $h(\mathbf{V}; \hat{\mathbf{V}})$ explicitly, and rearrange terms to obtain:

$$\begin{aligned}
h(\mathbf{V}; \hat{\mathbf{V}}) &= \sum_{i_k \in \mathcal{I}} \left(\tilde{a}_{i_k} + \hat{c}_{i_k} \text{Tr} \left[(\hat{\mathbf{E}}_{i_k})^{-1} \left(\hat{\mathbf{U}}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} + \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} - \hat{\mathbf{U}}_{i_k}^H \left(\sum_{(\ell,j)} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j_\ell} (\mathbf{V}_{j_\ell})^H (\mathbf{H}_{i_k}^\ell)^H \right) \hat{\mathbf{U}}_{i_k} \right) \right] \right) \\
&= \sum_{i_k \in \mathcal{I}} \left(\tilde{a}_{i_k} + \text{Tr} \left[\hat{c}_{i_k} (\hat{\mathbf{E}}_{i_k})^{-1} \left(\hat{\mathbf{U}}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} + \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} \right) - \underbrace{\sum_{(\ell,j)} \hat{c}_{j_\ell} (\mathbf{V}_{i_k})^H (\mathbf{H}_{j_\ell}^k)^H \hat{\mathbf{U}}_{j_\ell} (\hat{\mathbf{E}}_{j_\ell})^{-1} \hat{\mathbf{U}}_{j_\ell}^H \mathbf{H}_{j_\ell}^k \mathbf{V}_{i_k}}_{\text{quadratic term, only related to } \mathbf{V}_{i_k}} \right] \right) \\
&= \sum_{i_k \in \mathcal{I}} \left(\tilde{a}_{i_k} + \text{Tr} \left[\hat{c}_{i_k} (\hat{\mathbf{E}}_{i_k})^{-1} \left(\hat{\mathbf{U}}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} + \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} \right) - (\mathbf{V}_{i_k})^H \hat{\mathbf{J}}^k \mathbf{V}_{i_k} \right] \right) \\
&\triangleq \sum_{i_k \in \mathcal{I}} g_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}}).
\end{aligned} \tag{21}$$

where we have defined

$$\hat{\mathbf{J}}^k \triangleq \sum_{j_l \in \mathcal{I}} \hat{c}_{j_l} (\mathbf{H}_{j_l}^k)^H \hat{\mathbf{U}}_{j_l} (\hat{\mathbf{E}}_{i_k})^{-1} \hat{\mathbf{U}}_{j_l}^H \mathbf{H}_{j_l}^k \in \mathbb{S}_+^{MQ_k}. \quad (22)$$

It is now clear that $-h(\mathbf{V}; \hat{\mathbf{V}})$ is a quadratic convex function with respect to \mathbf{V} , hence the concavity of $h(\mathbf{V}; \hat{\mathbf{V}})$. Surprisingly, $h(\mathbf{V}; \hat{\mathbf{V}})$ can be written as a sum of the functions $\{g_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}})\}_{i_k \in \mathcal{I}}$, each of which is only related to a *single* variable in the set $\{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}}$. This interesting observation suggests that our series of approximations not only convexify the objective function $u(\mathbf{V})$, but more importantly *decompose* the non-linearly coupled objective function $u(\mathbf{V})$ into a sum of functions that decouple across the variables. This property of the sum-utility function $u(\mathbf{V})$ is one of the main results of this work, and will be leveraged for designing efficient distributed algorithms.

B. A General Successive Convex Approximation Algorithm

The lower bounds developed in the previous subsection are crucial for our design of efficient low-complexity algorithms that optimize the original problem (SYSTEM). In this subsection, we will develop the algorithm in its most general form. In the sections that follow, we will see how this algorithm can be effectively implemented and tailored for special cases of (SYSTEM).

Our approach is to successively approximate $f(\mathbf{V})$ using $h(\mathbf{V}; \hat{\mathbf{V}})$ to obtain progressively improved solutions. Let us use (t) to denote the iteration index. The proposed algorithm, referred to as the *successive convex approximation* (SCA) algorithm, can be carried out in the following three main steps.

Step 1: Suppose $\mathbf{V}(t-1)$ is a feasible solution to (SYSTEM). In iteration t , we solve the following convex optimization problem to obtain $\mathbf{V}(t)$

$$\begin{aligned} \max_{\mathbf{V}} \quad & h(\mathbf{V}; \mathbf{V}(t-1)) - s(\mathbf{V}) & (\text{Lower-Bound}) \\ \text{s.t.} \quad & \mathbf{V}^{q_k} \in \mathcal{V}^{q_k}, \forall q_k \in \mathcal{Q} \\ & \mathbf{V}^k \in \mathcal{V}^k, \forall k \in \mathcal{K}. \end{aligned}$$

Step 2: For each user $i_k \in \mathcal{I}$, compute

$$\mathbf{C}_{i_k}(t) = \sum_{(\ell, j)} \mathbf{H}_{i_k}^\ell \mathbf{V}_{j_\ell}(t) \mathbf{V}_{j_\ell}^H(t) (\mathbf{H}_{i_k}^\ell)^H + \sigma_{i_k}^2 \mathbf{I}_N \quad (24a)$$

$$\mathbf{U}_{i_k}(t) = (\mathbf{C}_{i_k}(t))^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k}(t) \quad (24b)$$

$$\mathbf{E}_{i_k}(t) = \mathbf{I}_{d_{i_k}} - \mathbf{V}_{i_k}^H(t) (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k}(t))^{-1} \mathbf{H}_{i_k}^k \mathbf{V}_{i_k}(t) \quad (24c)$$

$$c_{i_k}(t) = \frac{\partial f_{i_k}(x)}{\partial x} \Big|_{x = -\log |\mathbf{E}_{i_k}(t)|}. \quad (24d)$$

Note that when $f(\mathbf{V})$ takes the form of the popular weighted sum-rate, $c_{i_k}(t)$ is always a constant and step (24d) does not need to be performed.

Step 3: Compute the updated lower bound function $h(\mathbf{V}; \mathbf{V}(t))$ according to (21). Let $t = t + 1$, and go to Step 1.

A graphical illustration of the proposed algorithm is depicted in Fig. 3. It is important to note that efficiently solving the subproblem (Lower-Bound) in Step 1 is the key for the low-complexity implementation of the entire algorithm. Although this subproblem is already convex and thus can be solved using general purpose solvers (e.g., CVX [61]), in practice it is always desirable to develop customized low-complexity algorithms, tailored for problems with specific structures.

In the following, we will show that the proposed SCA algorithm converges to the set of stationary solutions of problem (SYSTEM). We note that the proof below differs slightly from a recent proof [62, Theorem 1], in which the convergence of a similar single-block successive approximation algorithm has been shown, with the variables assumed to be *real vectors*. Nevertheless, the main proof technique is the same as that of [62, Theorem 1], consequently the proof is included only for completeness.

To this end, the following definitions are needed:

- **Directional derivative:** Let $f : \mathcal{V} \rightarrow \mathbb{R}$ be a function where \mathcal{V} is a convex set. The directional derivative of f at point $x \in \mathcal{V}$ in direction d is defined by

$$f'_d(x) \triangleq \liminf_{r \downarrow 0} \frac{f(x + rd) - f(x)}{r}.$$

- **Stationary points of a function:** Let $f : \mathcal{V} \rightarrow \mathbb{R}$ where \mathcal{V} is a convex set. The point x is a stationary point of $f(\cdot)$ if $f'_d(x) \leq 0$ for all d such that $x + d \in \mathcal{V}$.

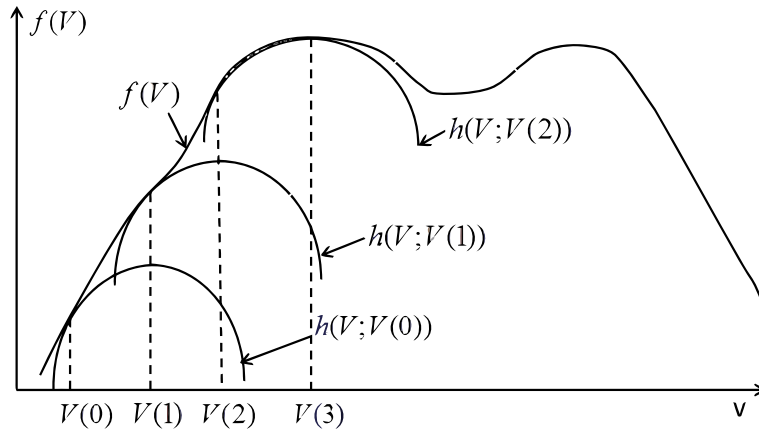


Fig. 3. A graphical illustration of the proposed SCA algorithm, assuming $s(\mathbf{V})$ is not present. At $\mathbf{V}(0)$, a concave function $h(\mathbf{V}; \mathbf{V}(0))$ is used to approximate the original non-convex function $f(\mathbf{V})$. The optimal solution of $h(\mathbf{V}; \mathbf{V}(0))$ is $\mathbf{V}(1)$. Then the concave function $h(\mathbf{V}; \mathbf{V}(1))$ is constructed at the point $\mathbf{V}(1)$, and optimized to obtain $\mathbf{V}(2)$. Continuing this process, a stationary solution of the original non-convex problem can be found.

Theorem 1 Suppose the following conditions hold:

- 1) Assumptions A-1)–A-3) and B-1)–B2) are all satisfied;
- 2) The convex problem (Lower-Bound) can be solved to its global optimality;

3) The sets $\{\mathcal{V}^k\}_{k \in \mathcal{K}}$ and $\{\mathcal{V}^{q_k}\}_{q_k \in \mathcal{Q}}$ are all convex, closed and compact.

Then the SCA algorithm converges to a stationary solution of the problem (SYSTEM) globally.

Proof: Firstly it is easy to observe that the objective value of the problem (SYSTEM) is monotonically nondecreasing, i.e., we have

$$u(\mathbf{V}(t+1)) \stackrel{(i)}{\geq} h(\mathbf{V}(t+1); \mathbf{V}(t)) - s(\mathbf{V}(t+1)) \stackrel{(ii)}{\geq} h(\mathbf{V}(t); \mathbf{V}(t)) - s(\mathbf{V}(t)) \stackrel{(iii)}{=} u(\mathbf{V}(t)) \quad (25)$$

where step (i) is from (20b); step (ii) is from the fact that $\mathbf{V}(t+1)$ is the solution to the problem (Lower-Bound); step (iii) is because of (20a). As both $f(\mathbf{V})$ and $s(\mathbf{V})$ are upper bounded for all \mathbf{V} in the feasible set, it follows that the sequence $\{u(\mathbf{V}(t))\}$ converges. Let \bar{u} denote its limit. This result combined with (25) implies that

$$\lim_{t \rightarrow \infty} h(\mathbf{V}(t+1); \mathbf{V}(t)) - s(\mathbf{V}(t+1)) = \bar{u}. \quad (26)$$

Using a similar argument as in [62, Lemma 1], and use the fact that $f(\cdot)$ and $h(\cdot; \hat{\mathbf{V}})$ satisfy (20a)–(20b), we can show that for any feasible $\hat{\mathbf{V}}$ the directional derivative of $f(\cdot)$ at the point $\hat{\mathbf{V}}$ equals that of $h(\cdot; \hat{\mathbf{V}})$ at the point $\hat{\mathbf{V}}$, i.e.,

$$\lim_{r \rightarrow 0} \frac{f(\hat{\mathbf{V}} + r\mathbf{D}) - f(\hat{\mathbf{V}})}{r} = \lim_{r \rightarrow 0} \frac{h(\hat{\mathbf{V}} + r\mathbf{D}; \hat{\mathbf{V}}) - h(\hat{\mathbf{V}})}{r} \quad (27)$$

where \mathbf{D} satisfies $\hat{\mathbf{V}} + \mathbf{D} \in \mathcal{V}$.

Let $\{\mathbf{V}(t_m)\}_{m=1}^\infty$ be a converging subsequence of $\mathbf{V}(t)$, and denote its limit as \mathbf{V}^* . In Step 1 of the algorithm, the optimality of $\mathbf{V}(t)$ to the problem (Lower-Bound) implies that

$$h(\mathbf{V}(t_m+1); \mathbf{V}(t_m)) - s(\mathbf{V}(t_m+1)) \geq h(\mathbf{V}; \mathbf{V}(t_m)) - s(\mathbf{V}), \forall \mathbf{V} \in \mathcal{V} \quad (28)$$

Taking limit for both sides and use (26), we obtain $\bar{u} \geq h(\mathbf{V}; \mathbf{V}^*) - s(\mathbf{V}), \forall \mathbf{V}$. Clearly we have

$$h(\mathbf{V}^*; \mathbf{V}^*) - s(\mathbf{V}^*) = u(\mathbf{V}^*) = \bar{u}.$$

This implies that $h(\mathbf{V}^*; \mathbf{V}^*) - s(\mathbf{V}^*) \geq h(\mathbf{V}; \mathbf{V}^*) - s(\mathbf{V})$ for all feasible \mathbf{V} , which further implies that

$$h'_{\mathbf{D}}(\mathbf{V}; \mathbf{V}^*) \Big|_{\mathbf{V}=\mathbf{V}^*} - s'_{\mathbf{D}}(\mathbf{V}) \Big|_{\mathbf{V}=\mathbf{V}^*} \leq 0, \forall \mathbf{D} + \mathbf{V}^* \in \mathcal{V}. \quad (29)$$

Utilizing (27), we obtain

$$u'_{\mathbf{D}}(\mathbf{V}) \Big|_{\mathbf{V}=\mathbf{V}^*} = f'_{\mathbf{D}}(\mathbf{V}) \Big|_{\mathbf{V}=\mathbf{V}^*} - s'_{\mathbf{D}}(\mathbf{V}) \Big|_{\mathbf{V}=\mathbf{V}^*} \leq 0, \forall \mathbf{D} + \mathbf{V}^* \in \mathcal{V}, \quad (30)$$

which says that \mathbf{V}^* is a stationary solution to the problem (SYSTEM). ■

IV. CUSTOMIZED ALGORITHMS FOR PRECODER DESIGN

In this section, we will customize the general SCA algorithm proposed in the previous section to the precoder design problems in different network setting. The main focus is to explore the structure of different problems so that the convex subproblem (Lower-Bound) can be solved efficiently for each network scenario.

A. Linear Precoder Design for IBC Model

First let us consider the IBC model in which in each cell, there is a single BS transmitting. The objective is to design linear precoder for the users subject to the sum-power constraint for each BS. As $|\mathcal{Q}_k| = 1$ for all k , \mathbf{V}_{i_k} denotes the transmit precoder used by the BS in cell k to transmit to user i_k . In this case, there is no penalty term $s(\mathbf{V})$, and the sets \mathcal{V}^{q_k} and \mathcal{V}^k collapse to a single feasible set given by $\mathcal{V}^k = \sum_{i_k \in \mathcal{I}_k} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq \bar{P}_k$. Hence the subproblem (Lower-Bound) is given by (expanding $h(\mathbf{V}; \hat{\mathbf{V}})$ using $\sum_{i \in \mathcal{I}} g_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}})$, cf. (21))

$$\begin{aligned} & \max_{\{\mathbf{V}^k\}_{k \in \mathcal{K}}} \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}}) \\ & \text{s.t.} \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq \bar{P}_k, \quad k \in \mathcal{K}. \end{aligned} \quad (\text{IBC})$$

Clearly both the constraints and the objective of the above problem are separable among the BSs (i.e., separable among the set of variables $\mathbf{V}^k = \{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}_k}$, for different $k \in \mathcal{K}$). As a result, we can decompose this problem into K independent subproblems, with the k -th subproblem taking the following form

$$\begin{aligned} & \max_{\mathbf{V}^k} g^k(\mathbf{V}^k; \hat{\mathbf{V}}) \\ & \text{s.t.} \quad g^k(\mathbf{V}^k; \hat{\mathbf{V}}) = \sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}, \hat{\mathbf{V}}) \\ & \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr}(\mathbf{V}_{i_k} \mathbf{V}_{i_k}^H) \leq \bar{P}_k. \end{aligned} \quad (\text{IBC-SUB})$$

Let $\lambda^k \geq 0$ denote the Lagrangian multiplier associated with the power constraint. Then the Lagrangian function for problem (IBC-SUB) can be expressed as (excluding constants that are not related to $(\lambda^k, \mathbf{V}^k)$)

$$\begin{aligned} L_k(\mathbf{V}^k, \lambda^k; \hat{\mathbf{V}}) = & \sum_{i_k \in \mathcal{I}_k} \left(\text{Tr} \left[\hat{\mathbf{c}}_{i_k} (\hat{\mathbf{E}}_{i_k})^{-1} \left(\hat{\mathbf{U}}_{i_k}^H \mathbf{H}_{i_k}^k \mathbf{V}_{i_k} + \mathbf{V}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} \right) \right. \right. \\ & \left. \left. - (\mathbf{V}_{i_k})^H \hat{\mathbf{J}}^k \mathbf{V}_{i_k} \right] - \lambda^k \left(\sum_{i_k \in \mathcal{I}_k} \text{Tr} [\mathbf{V}_{i_k} (\mathbf{V}_{i_k})^H] - \bar{P}_k \right) \right). \end{aligned} \quad (31)$$

The dual function is $d(\lambda^k; \hat{\mathbf{V}}) = \max_{\mathbf{V}^k} L(\mathbf{V}^k, \lambda^k; \hat{\mathbf{V}})$. The optimal primal-dual pair $((\mathbf{V}^k)^*, (\lambda^k)^*)$ should satisfy the KKT optimality conditions

$$(\mathbf{V}^k)^* = \arg \max_{\mathbf{V}^k} L_k(\mathbf{V}^k, (\lambda^k)^*; \hat{\mathbf{V}}), \quad (32a)$$

$$0 \leq (\lambda^k)^* \perp \bar{P}_k - \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^* (\mathbf{V}_{i_k}^*)^H] \geq 0. \quad (32b)$$

For fixed $\lambda^k \geq 0$, the solution for the unconstrained problem $\max_{\mathbf{V}^k} L(\mathbf{V}^k, \lambda^k; \hat{\mathbf{V}})$, denoted as $\{\mathbf{V}_{i_k}^*(\lambda^k)\}_{i_k \in \mathcal{I}_k}$, can be expressed as

$$\begin{aligned} \mathbf{V}_{i_k}^*(\lambda^k) &= \left(\hat{\mathbf{J}}^k + \lambda^k \mathbf{I}_M \right)^{-1} \hat{\mathbf{c}}_{i_k}(\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k}(\hat{\mathbf{E}}_{i_k})^{-1} \\ &= \left(\sum_{(\ell, j)} \hat{\mathbf{c}}_{j\ell}(\mathbf{H}_{j\ell}^k)^H \hat{\mathbf{U}}_{\ell j}(\hat{\mathbf{E}}_{j\ell})^{-1} \hat{\mathbf{U}}_{\ell j}^H \mathbf{H}_{j\ell}^k + \lambda^k \mathbf{I}_M \right)^{-1} \hat{\mathbf{c}}_{i_k}(\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k}(\hat{\mathbf{E}}_{i_k})^{-1}, \quad \forall i_k \in \mathcal{I}_k. \end{aligned} \quad (33)$$

To find the optimal multiplier $(\lambda^k)^*$ that satisfies the complementarity condition (32b), we utilize a general result on penalty method for optimization, e.g., [63, Section 12.1, Lemma 1]. This result asserts that for the solution $\{\mathbf{V}_{i_k}^*(\lambda^k)\}_{i_k \in \mathcal{I}_k}$, the penalized term $\sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^*(\lambda^k)(\mathbf{V}_{i_k}^*(\lambda^k))^H]$ must be monotonically decreasing with respect to λ^k . Such monotonicity result suggests that we can find the optimal multiplier that satisfies (32b) by a simple bisection search procedure.

The algorithm discussed in this subsection is summarized in Table II. The small constant $\delta \geq 0$ in Step S4) is used to specify the stopping criterion.

TABLE II
THE PROPOSED ALGORITHM FOR (SYSTEM) IN IBC SETTING

<p>S1): Initialization Obtain a feasible solution $\mathbf{V}_{i_k}(0)$ for all $i_k \in \mathcal{I}$</p> <p>S2): For each BS k, compute $\mathbf{C}_{i_k}(t)$, $\mathbf{E}_{i_k}(t)$, $\mathbf{U}_{i_k}(t)$ and $c_{i_k}(t)$, according to (24a)–(24d), for all $i_k \in \mathcal{I}_k$</p> <p>S3): For each BS k, compute the precoders $\mathbf{V}^k(t)$ by</p> $\mathbf{V}_{i_k}(t) = \left(\sum_{(\ell, j)} c_{j\ell}(t)(\mathbf{H}_{j\ell}^k)^H \mathbf{U}_{\ell j}(t)(\mathbf{E}_{j\ell}(t+1))^{-1} \mathbf{U}_{\ell j}^H(t) \mathbf{H}_{j\ell}^k + (\lambda^k)^* \mathbf{I}_M \right)^{-1} \times c_{i_k}(t)(\mathbf{H}_{i_k}^k)^H \mathbf{U}_{i_k}(t)(\mathbf{E}_{i_k}(t))^{-1}, \quad \forall i_k \in \mathcal{I}_k$ <p>where $(\lambda^k)^*$ is computed by a bisection procedure</p> <p>S4) Until some stopping criterion criterion is met</p>

Remark 1 (*The bisection procedure*): The computation for each bisection step for finding the optimal multiplier $(\lambda^k)^*$ can be carried out as follows [44]: 1) Perform an eigen-decomposition $\sum_{(\ell, j)} \hat{\mathbf{c}}_{j\ell}(\mathbf{H}_{j\ell}^k)^H \hat{\mathbf{U}}_{\ell j}(\hat{\mathbf{E}}_{j\ell})^{-1} \hat{\mathbf{U}}_{\ell j}^H \mathbf{H}_{j\ell}^k = \mathbf{X}_k \Phi_k \mathbf{X}_k^H$, where Φ_k is a diagonal matrix; 2) For a given $\lambda_k \geq 0$, and for each $i_k \in \mathcal{I}_k$, compute $\mathbf{V}_{i_k}^*(\lambda^k) = \mathbf{X}_k(\Phi_k^{-1} + \lambda^k \mathbf{I}) \mathbf{X}_k^H \hat{\mathbf{c}}_{i_k}(\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k}(\hat{\mathbf{E}}_{i_k})^{-1}$. In this way, there is no need to compute the matrix inversion in each step of the bisection.

Remark 2 (*Relationship with WMMSE algorithm*): The algorithm listed in Table II recovers the weighted-MMSE (WMMSE) algorithm proposed recently in [44] for utility maximization problem in IBC networks¹. The original WMMSE algorithm is derived base on ceratin equivalence relationship between the sum utility optimization problem and a weighted MMSE minimization problem (see [44, Section II-A]), while in this paper we arrive at this algorithm

¹Although there are several subtle differences between these two algorithms. For example, proving convergence of the WMMSE algorithm requires that $f_{i_k}(-\log |X|)$ is strictly convex on its argument, and that the subproblem (IBC) admits a unique solution. There are no such requirements for the SCA-based algorithm.

by specializing the SCA algorithm to the IBC setting. Thus the SCA algorithm is more general and includes the WMMSE algorithm as a special case.

Remark 3 (*Precoder Design for HetNet with both intra-cell and inter-cell CB*): The algorithm in Table II is also applicable for the linear precoder design problem in HetNet with *both* intra-cell and inter-cell CB. In this case, the BSs within each cell only cooperate in the precoder level, and each user is served by a single BS. In this case, the entire network can be viewed as an IBC with $K \times |\mathcal{Q}_k|$ BSs, thus the algorithm in Table II can be directly applied.

B. IBC Model Intra-cell ZF and Inter-cell CB Design

We then consider a similar IBC model as in the previous subsection, but with each BS employing a ZF precoder to cancel the intra-cell interference among the users. We assume that certain user-selection within each cell has already been performed, which ensures that the per-cell zero forcing constraint $\mathbf{H}_{j_k,k} \mathbf{V}_{i_k} (\mathbf{V}_{i_k})^H \mathbf{H}_{j_k,k}^H = \mathbf{0}, \forall j_k \neq i_k, j_k \in \mathcal{I}_k$ is always feasible (one easily checkable condition that guarantees feasibility is $|\mathcal{I}_k| \times N \leq M$, see [19], [20]). Note that in this network setting, although the intra-cell interference is canceled by the use of ZF precoder, the inter-cell interference is still present. Hence the original problem (SYSTEM) is still difficult to solve. The subproblem (Lower-Bound) can be specialized to have the following form

$$\max_{\mathbf{V}} \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}}) \quad (\text{IBC-ZF1})$$

$$\text{s.t.} \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr} [\mathbf{V}_{i_k} (\mathbf{V}_{i_k})^H] \leq \bar{P}_k, \quad k \in \mathcal{K} \quad (34a)$$

$$\mathbf{H}_{j_k,k} \mathbf{V}_{i_k} (\mathbf{V}_{i_k})^H \mathbf{H}_{j_k,k}^H = \mathbf{0}, \forall j_k \neq i_k, j_k \in \mathcal{I}_k, k \in \mathcal{K}. \quad (34b)$$

To remove the ZF constraints (34b), let $L \triangleq N(|\mathcal{I}_k| - 1)$, and define a set of concatenated channel matrices

$$\mathbf{Q}_{i_k} \triangleq \{\mathbf{H}_{j_k}\}_{j_k \in \mathcal{I}_k \setminus \{i_k\}} \in \mathbb{C}^{L \times M}, \quad \forall i_k \in \mathcal{I}_k. \quad (35)$$

Let $\mathbf{Q}_{i_k} = \mathbf{L}_{i_k} \mathbf{\Sigma}_{i_k} \mathbf{R}_{i_k}$ denote the singular value decomposition of \mathbf{Q}_{i_k} , where \mathbf{L}_{i_k} and \mathbf{R}_{i_k} are two unitary matrices, with $\mathbf{L}_{i_k} \in \mathbb{C}^{L \times L}$ and $\mathbf{R}_{i_k} \in \mathbb{C}^{L \times M}$, and $\mathbf{\Sigma}_{i_k}$ being an $L \times L$ diagonal matrix. Let $\mathbf{P}_{i_k} = (\mathbf{I} - \mathbf{R}_{i_k} \mathbf{R}_{i_k}^H)$ denote a projection matrix to the space orthogonal to that spanned by \mathbf{R}_{i_k} . Let $\mathbf{P}_{i_k} = \tilde{\mathbf{R}}_{i_k} \tilde{\mathbf{R}}_{i_k}^H$, where $\tilde{\mathbf{R}}_{i_k} \in \mathbb{C}^{M \times (M-L)}$ is composed of the orthogonal basis that satisfies $\mathbf{R}_{i_k} \tilde{\mathbf{R}}_{i_k} = \mathbf{0}$ and $\tilde{\mathbf{R}}_{i_k}^H \tilde{\mathbf{R}}_{i_k} = \mathbf{I}$. Then [21, Lemma 3.1] asserts ² that the optimal solution of problem (IBC-ZF1) must be of the form: $\mathbf{V}_{i_k} = \tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k}$, where $\mathbf{W}_{i_k} \in \mathbb{C}^{(M-L) \times d_{i_k}}$.

²Note that Lemma 3.1 in [21] still applies in our setting as the property derived in that lemma is *not* related to the form of the objective function. It is only related to the zero-forcing constraints (34b).

Utilizing this structure of the optimal solution, the problem (IBC-ZF1) can be equivalently written by

$$\begin{aligned} \max_{\mathbf{W}} \quad & \sum_{k \in \mathcal{K}} \sum_{i_k \in \mathcal{I}_k} \tilde{g}_{i_k}(\mathbf{W}_{i_k}; \hat{\mathbf{V}}) \\ \text{s.t.} \quad & \sum_{i_k \in \mathcal{I}_k} \text{Tr}(\tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k} (\tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k})^H) \leq \bar{P}_k, \quad k \in \mathcal{K}, \end{aligned} \quad (\text{IBC-ZF2})$$

where the function $\tilde{g}_{i_k}(\cdot)$ is the same as the original objective $g_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}})$, except that \mathbf{V}_{i_k} is replaced by $\tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k}$:

$$\begin{aligned} \tilde{g}_{i_k}(\mathbf{W}_{i_k}; \hat{\mathbf{V}}) \triangleq & \tilde{a}_{i_k} + \text{Tr} \left[\hat{c}_{i_k} (\hat{\mathbf{E}}_{i_k})^{-1} \left(\hat{\mathbf{U}}_{i_k}^H \mathbf{H}_{i_k}^k \tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k} + (\tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k})^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} \right) \right. \\ & \left. - \sum_{(\ell, j)} \hat{c}_{j\ell} (\tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k})^H (\mathbf{H}_{j\ell}^k)^H \hat{\mathbf{U}}_{\ell j} (\hat{\mathbf{E}}_{j\ell})^{-1} \hat{\mathbf{U}}_{\ell j}^H \mathbf{H}_{j\ell}^k \tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k} \right] \end{aligned} \quad (37)$$

Again both the constraints and the objective of this problem are separable among the BSs (i.e., the set of variables $\{\mathbf{V}_{i_k}\}_{i_k \in \mathcal{I}_k}$), and we can further decompose this problem into K independent subproblems of the form

$$\begin{aligned} \max_{\mathbf{W}^k} \quad & \tilde{g}^k(\mathbf{W}^k; \hat{\mathbf{V}}) \\ \text{s.t.} \quad & \tilde{g}^k(\mathbf{W}^k; \hat{\mathbf{V}}) = \sum_{i_k \in \mathcal{I}_k} \tilde{g}_{i_k}(\mathbf{W}_{i_k}; \hat{\mathbf{V}}) \\ & \sum_{i_k \in \mathcal{I}_k} \text{Tr}(\tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k} (\tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k})^H) \leq \bar{P}_k, \end{aligned} \quad (\text{IBC-ZF-SUB})$$

Let us use λ_k to denote the Lagrangian multiplier associated with the power constraint. Then by using similar steps that lead to (33), we can show that the optimal solution for problem (IBC-ZF-SUB) is of the form

$$\mathbf{W}_{i_k}^* = \left(\sum_{(\ell, j)} \hat{c}_{j\ell} (\mathbf{H}_{j\ell}^k \tilde{\mathbf{R}}_{i_k})^H \hat{\mathbf{U}}_{\ell j} (\hat{\mathbf{E}}_{j\ell})^{-1} \hat{\mathbf{U}}_{\ell j}^H \mathbf{H}_{j\ell}^k \tilde{\mathbf{R}}_{i_k} + \lambda_k^* \mathbf{I}_{M-L} \right)^{-1} \hat{c}_{i_k} \tilde{\mathbf{R}}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} (\hat{\mathbf{E}}_{i_k})^{-1}, \quad \forall i_k \in \mathcal{I}_k. \quad (39)$$

where λ_k^* can be computed by a bisection method.

The algorithm discussed in this section is summarized in Table III.

TABLE III
THE PROPOSED ALGORITHM FOR (SYSTEM) IN IBC SETTING WITH INTRA-CELL ZF AND INTER-CELL CB

S1): Initialization

S1a): For each BS k , compute:

$$\mathbf{Q}_{i_k} = \mathbf{I}_{i_k} \Sigma_{i_k} \mathbf{R}_{i_k}, \mathbf{P}_{i_k} = (\mathbf{I} - \mathbf{R}_{i_k} \mathbf{R}_{i_k}^H), \text{ and } \mathbf{P}_{i_k} = \tilde{\mathbf{R}}_{i_k} \tilde{\mathbf{R}}_{i_k}^H, \quad \forall i_k \in \mathcal{I}_k$$

S1b): Obtain a feasible solution $\mathbf{V}_{i_k}(0)$ for all $i_k \in \mathcal{I}$

S2): For each BS k , compute $\mathbf{C}_{i_k}(t)$, $\mathbf{E}_{i_k}(t)$, $\mathbf{U}_{i_k}(t)$ and $c_{i_k}(t)$, according to (24a)–(24d), for all $i_k \in \mathcal{I}_k$

$$\begin{aligned} \text{S3a): } \mathbf{W}_{i_k}(t) = & \left(\sum_{(\ell, j)} c_{j\ell}(t) (\tilde{\mathbf{R}}_{i_k} \mathbf{H}_{j\ell}^k)^H \mathbf{U}_{\ell j}(t) (\mathbf{E}_{j\ell}(t))^{-1} (\mathbf{U}_{\ell j}(t))^H \mathbf{H}_{j\ell}^k \tilde{\mathbf{R}}_{i_k} + \lambda_k^* \mathbf{I}_{M-L} \right)^{-1} \\ & \times c_{i_k}(t) \tilde{\mathbf{R}}_{i_k}^H (\mathbf{H}_{i_k}^k)^H \mathbf{U}_{i_k}(t) (\mathbf{E}_{i_k}(t))^{-1}, \quad \forall i_k \in \mathcal{I}_k \end{aligned}$$

where λ_k^* is computed by a bisection procedure

$$\text{S3b): } \mathbf{V}_{i_k}(t) = \tilde{\mathbf{R}}_{i_k} \mathbf{W}_{i_k}(t)$$

S4) Until some stopping criterion is met.

C. Linear Precoder Design for HetNet with Intra-Cell Full ComP and Inter-Cell CB

Consider a HetNet setting, in which there are a set of \mathcal{Q}_k BSs in each cell, and they form a single *virtual* BS to transmit to the users. The objective is to design the virtual linear precoder for the users subject to the sum-power constraint for each individual BS. In this case, \mathcal{V}^{q_k} becomes $\mathcal{V}^{q_k} = \{\mathbf{V}^{q_k} : \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^{q_k} (\mathbf{V}_{i_k}^{q_k})^H] \leq \bar{P}^{q_k}\}$. Assume for now that there is no penalty term $s(\mathbf{V})$. Then the subproblem (Lower-Bound) can be again decomposed into K independent subproblems of the form

$$\begin{aligned} \max_{\mathbf{V}_k} \quad & \sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}, \hat{\mathbf{V}}) \\ \text{s.t.} \quad & \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^{q_k} (\mathbf{V}_{i_k}^{q_k})^H] \leq \bar{P}^{q_k}, \quad \forall q_k \in \mathcal{Q}_k. \end{aligned} \quad (\text{VIBC-SUB})$$

Differently from problem (IBC-SUB) discussed in Section IV-A, the above problem has $Q_k = |\mathcal{Q}_k|$ separable constraints (each constraining a subset of variables), hence Q_k Lagrangian multipliers $\{\lambda_k^{q_k}\}_{q_k \in \mathcal{Q}_k}$. The bisection algorithm on a single multiplier developed in Table II thus does not work in this case.

Fortunately, the constraints for this problem are separable among different block variables $\{\mathbf{V}^{q_k}\}_{q_k \in \mathcal{Q}_k}$ (that is, precoders belongs to different BSs). Therefore a natural way to obtain the optimal solution for the problem (VIBC-SUB) is to use a block coordinate descent (BCD) algorithm (see [64], [65]), which updates one block variable \mathbf{V}^{q_k} at a time while holding the remaining block variables fixed. To capitalize the block structure of the objective of problem (VIBC-SUB), the following definitions are needed. Let

$$\hat{\mathbf{S}}_{i_k} \triangleq \hat{c}_{i_k} (\mathbf{H}_{i_k}^k)^H \hat{\mathbf{U}}_{i_k} (\hat{\mathbf{E}}_{i_k})^{-1} \in \mathbb{C}^{M \times Q_k \times d_{i_k}}, \quad \forall i_k \in \mathcal{I}_k. \quad (40)$$

Partition $\hat{\mathbf{J}}^k$ (which is defined in (22)) and $\hat{\mathbf{S}}_{i_k}$ into the following form

$$\hat{\mathbf{J}}^k = \begin{bmatrix} \hat{\mathbf{J}}^k[1, 1], & \cdots, & \hat{\mathbf{J}}^k[1, Q_k] \\ \vdots & \ddots & \vdots \\ \hat{\mathbf{J}}^k[Q_k, 1] & \cdots & \hat{\mathbf{J}}^k[Q_k, Q_k] \end{bmatrix}, \quad \hat{\mathbf{S}}_{i_k} = [\hat{\mathbf{S}}_{i_k}^H[1], \dots, \hat{\mathbf{S}}_{i_k}^H[Q_k]]^H \quad (41)$$

where $\hat{\mathbf{J}}^k[q, p] \in \mathbb{C}^{M \times M}$, $\forall (q, p) \in \mathcal{Q}_k \times \mathcal{Q}_k$, and $\hat{\mathbf{S}}_{i_k}[q] \in \mathbb{C}^{M \times d_{i_k}}$, $\forall q \in \mathcal{Q}_k$. Then the function $g_{i_k}(\mathbf{V}_{i_k}, \hat{\mathbf{V}})$ defined in (21) can be expressed as

$$g_{i_k}(\mathbf{V}_{i_k}, \hat{\mathbf{V}}) = \tilde{a}_{i_k} + \sum_{p_k \in \mathcal{Q}_k} \text{Tr} \left[\hat{\mathbf{S}}_{i_k}^H[p_k] \mathbf{V}_{i_k}^{p_k} + (\mathbf{V}_{i_k}^{p_k})^H \hat{\mathbf{S}}_{i_k}[p_k] \right] - \sum_{p_k, q_k \in \mathcal{Q}_k} \text{Tr} \left[(\mathbf{V}_{i_k}^{q_k})^H \hat{\mathbf{J}}^k[q_k, p_k] \mathbf{V}_{i_k}^{p_k} \right] \quad (42)$$

Clearly, $\sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}, \hat{\mathbf{V}})$ is again a quadratic function with respect to one particular block variable, say \mathbf{V}^{m_k} .

It follows that the per-block problem, written in the following form, can be efficiently solved in closed form.

$$\begin{aligned}
 & \max_{\mathbf{V}^{m_k}} \quad g^k(\mathbf{V}^k, \hat{\mathbf{V}}) \\
 & \text{s.t.} \quad g^k(\mathbf{V}^k, \hat{\mathbf{V}}) = \sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}, \hat{\mathbf{V}}), \\
 & \quad \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^{m_k} (\mathbf{V}_{i_k}^{m_k})^H] \leq \bar{P}^{m_k}.
 \end{aligned} \tag{VIBC-BLK}$$

Let $\lambda^{m_k} \geq 0$ denote the Lagrangian multiplier associated with the power constraint of the m_k -th subproblem. Following the same derivation in Section IV-A, the optimal solution $(\mathbf{V}^{m_k})^*$ for the problem (VIBC-BLK) can be expressed as

$$(\mathbf{V}_{i_k}^{m_k})^* = \left(\hat{\mathbf{J}}^k[m_k, m_k] + (\lambda^{m_k})^* \mathbf{I}_M \right)^{-1} \left(\hat{\mathbf{S}}_{i_k}[m_k] - \sum_{p_k \neq m_k} \hat{\mathbf{J}}^k[m_k, p_k] \mathbf{V}_{i_k}^{p_k} \right), \quad \forall i_k \in \mathcal{I}_k \tag{43}$$

where the optimal multiplier can be computed again using a bisection search.

In summary, in the presence of multiple BSs with individual power constraint, the proposed algorithm consists of the following two layers: *i)* the outer layer that updates $\mathbf{E}_{i_k}(t)$, $\mathbf{U}_{i_k}(t)$, $c_{i_k}(t)$, $\mathbf{J}^k(t)$ and $\mathbf{S}^k(t)$; *ii)* the inner layer that updates each \mathbf{V}^k by a BCD algorithm with blocks given by $\{\mathbf{V}^{q_k}\}_{q_k \in \mathcal{Q}_k}$. The overall algorithm is detailed in Table IV.

TABLE IV
THE PROPOSED ALGORITHM FOR SOLVING (SYSTEM) WITH INTRA-CELL COMP AND INTER-CELL CB

S1): **Initialization** Obtain a feasible solution $\mathbf{V}_{i_k}(0)$ for all $i_k \in \mathcal{I}$
S2): For each BS k , compute $\mathbf{C}_{i_k}(t)$, $\mathbf{E}_{i_k}(t)$, $\mathbf{U}_{i_k}(t)$ and $c_{i_k}(t)$, according to (24a)–(24d), for all $i_k \in \mathcal{I}_k$
S3): For each BS k , compute the following
S3a): $\mathbf{J}^k(t) = \sum_{j_l \in \mathcal{I}} c_{j_l}(t) (\mathbf{H}_{j_l}^k)^H \mathbf{U}_{j_l}(t) (\mathbf{E}_{i_k}(t))^{-1} (\mathbf{U}_{j_l}(t))^H \mathbf{H}_{j_l}^k$
S3b): $\mathbf{S}_{i_k}(t) = c_{i_k}(t) (\mathbf{H}_{i_k}^k)^H \mathbf{U}_{i_k}(t) (\mathbf{E}_{i_k}(t))^{-1}$, $\forall i_k \in \mathcal{I}_k$.
S4): For each BS k , compute the precoders $\mathbf{V}^k(t)$ by
Repeat Cyclically pick $m_k \in \mathcal{Q}_k$
 Compute $(\mathbf{V}_{i_k}^{m_k})^*$ using (43), $\forall i_k \in \mathcal{I}_k$
 where $(\lambda^{m_k})^*$ is computed by a bisection procedure
Until Desired stopping criterion is met
 Let $\mathbf{V}_{i_k}^{m_k}(t) = (\mathbf{V}_{i_k}^{m_k})^*$, $\forall i_k, m_k$
S5) Until some stopping criteria is met.

Remark 4 The convergence of the BCD step S4) to the global optimal solution of the subproblem (VIBC-BLK) can be shown using standard argument such as [65, Theorem 4.1].

Remark 5 (Hybrid Implementation) Interestingly, the proposed framework further allows for the implementation of *hybrid* cooperation schemes, in which the cells can choose to serve the users by using either ZF based precoding or the general linear precoding. Moreover, certain cells can have only a single BS, while the rest can have multiple of them. Such hybrid implementation amounts to requiring each subproblem (Lower-Bound) take different forms

of constraints. As long as each subproblem can be solved to its global optimality, the convergence of the overall algorithm is always guaranteed.

Remark 6 (*Per-cell partial ComP*): The algorithm proposed in this section can be easily extended to the case of per-cell partial ComP.

Assume that the BS clustering structure is known ³, and letting $\mathcal{S}^{q_k} \subseteq \mathcal{I}_k$ denote the set of users served by BS q_k , then we only need to slightly modify the proposed algorithm in Table IV by the following:

- 1) In S1), for each BS $m_k \in \mathcal{Q}_k$, set $\mathbf{V}_{i_k}^{m_k}(0) = \mathbf{0}$ for all $i_k \notin \mathcal{S}^{m_k}$; find a feasible solution for the rest of users $i_k \in \mathcal{S}^{m_k}$;
- 2) In S4), let each BS $m_k \in \mathcal{Q}_k$ compute $\mathbf{V}_{i_k}^{m_k}(t)$ using (43), $\forall i_k \in \mathcal{S}^{m_k}$ (instead of for all $i_k \in \mathcal{I}_k$).

In this way, only the precoders of the subset of users served by each BS will be updated in each iteration. Once again, the computation in each iteration admits a closed-form solution, while in related works such as [53], general purpose convex solvers need to be used for solving the subproblems ⁴.

Moreover, when the BSs' clustering structure needs to be designed jointly with the precoders, we can include the penalty term $s(\mathbf{V})$ into the objective to induce certain block-sparsity in the precoder \mathbf{V}^{q_k} . Except for this additional term in the objective, which leads to different solution to the associated subproblem, the algorithm for the joint BS clustering and precoder design problem is the same as the one in Table IV. Specifically, the per-block subproblem (VIBC-BLK) takes the following form

$$\begin{aligned} \max_{\mathbf{V}^{m_k}} \quad & g^k(\mathbf{V}^k, \hat{\mathbf{V}}) - \sum_{i_k \in \mathcal{I}_k} s_{i_k}^{m_k}(\mathbf{V}_{i_k}^{m_k}) \\ \text{s.t.} \quad & g^k(\mathbf{V}^k, \hat{\mathbf{V}}) = \sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}, \hat{\mathbf{V}}), \\ & \sum_{i_k \in \mathcal{I}_k} \text{Tr}[\mathbf{V}_{i_k}^{m_k} (\mathbf{V}_{i_k}^{m_k})^H] \leq \bar{P}^{m_k}. \end{aligned} \tag{44}$$

In particular, when we let $s_{i_k}^{m_k}(\mathbf{V}_{i_k}^{m_k}) = \|\mathbf{V}_{i_k}^{m_k}\|$, this subproblem becomes a well known quadratic group-LASSO problem [66] (with an additional quadratic constraint), which can be solved using an iterative procedure. We refer the readers to [54] for detailed algorithm ⁵.

The algorithm proposed in Table IV can be viewed as a *double time-scale* algorithm: the subproblem (VIBC-SUB) is solve in a relatively fast time-scale by a BCD iteration (i.e., step S4)), while the computation in S2)-S3) is performed less frequently in a slow time-scale.

³Such clustering structure can be determined, for example, by using simple heuristic pathloss model. A particular useful scheme for BS clustering is to serve the users by the sets of BSs that are adjacent to them. where the closeness of the BSs to the users are measured by the pathloss coefficients among the users and BSs.

⁴The subproblem used in [53] is derived from certain difference of convex function (d.c.) property of the weighted sum-rate objective, which is, of course, in a very different form from the one used in the present work.

⁵Reference [54] solves a slightly simplified *beamforming* problem when $N = 1$ and $d_{i_k} = 1$ for all $i_k \in \mathcal{I}_k$. That is, there is a *single* data stream for each user. Nevertheless, the algorithm therein can be extended easily to the more general case with $N > 1$ and $d_{i_k} > 1$.

One desirable feature offered by this separation in time-scale is that in each of its fast time scales, interactions are only limited within each cell: there is no coordination or message exchanges required among the BSs in different cells. On the other hand, however, to guarantee the convergence of the overall algorithm, the fast time-scale computation needs to be performed *until* the subproblem (VIBC-SUB) is solved *exactly*. Such requirement turns out to be quite inflexible for practical implementation. The main reasons are listed below:

- 1) It is usually difficult to check whether the inner iteration has indeed reached the optimality;
- 2) Before reaching the optimality for the subproblem (VIBC-SUB), the marginal benefit of the precoder updates in the inner iteration decreases as the iteration progresses. This effect is manifested in particular in the first few outer iterations, in which even the inner problem is solved exactly, the precoders obtained are still far away from the optimal ones.

Heuristically, one may resort to running a few, or even a single, BCD iterations for each cell in Step S4), but the convergence properties of such heuristics seem hard to establish. Fortunately, the universal lower bound established in Section III allows us to develop a different version of the SCA algorithm, in which each subproblem (Lower-Bound) is solved *inexactly*. The benefit of such algorithm is quite obvious from our preceding discussion. It allows one to solve the subproblems approximately at the beginning, and more accurately later as the iteration progresses. In the next section, we will present in detail the “inexact” version of the SCA algorithm, analyze its convergence properties, and demonstrate its possible application in the precoder design problem in the HetNet.

V. AN EXTENDED INEXACT SCA APPROACH

In this section, we present an important extension of the SCA algorithm, which is a *single* time-scale algorithm *without* any inner iterations. Once again, such single time-scale implementation of the SCA algorithm hinges upon our ability to decompose the original problem (SYSTEM) into a sequence of separable and convex subproblems in the form of (Lower-Bound). The key difference of the inexact version of the SCA algorithm, compared with its original counterpart, is that in each of its iteration, a reasonably “good” step that improves the objective of problem (Lower-Bound) is taken to update the precoder \mathbf{V} , instead of finding the one that solves the subproblem (Lower-Bound) to its global optimality.

For clarity of presentation, we will first introduce the algorithm in its general form and prove its convergence. As an example, we will then specialize it for the HetNet linear transceiver design problem studied in Section IV-C.

A. The Inexact-SCA Algorithm

For simplicity of notation, let us introduce the following definitions:

$$g^k(\mathbf{V}^k; \hat{\mathbf{V}}) \triangleq \sum_{i_k \in \mathcal{I}_k} g_{i_k}(\mathbf{V}_{i_k}; \hat{\mathbf{V}}) \quad (45a)$$

$$s^k(\mathbf{V}^k) \triangleq \sum_{i_k \in \mathcal{I}_k} \sum_{q_k \in \mathcal{Q}_k} s_{i_k}^{q_k}(\mathbf{V}_{i_k}^{q_k}) \quad (45b)$$

$$u^k(\mathbf{V}^k; \hat{\mathbf{V}}) \triangleq g^k(\mathbf{V}^k; \hat{\mathbf{V}}) - s^k(\mathbf{V}^k). \quad (45c)$$

Using these definitions, the overall system lower bound $h(\mathbf{V}; \hat{\mathbf{V}}) - s(\mathbf{V})$ can be expressed as

$$h(\mathbf{V}; \hat{\mathbf{V}}) - s(\mathbf{V}) = \sum_{k \in \mathcal{K}} u^k(\mathbf{V}^k; \hat{\mathbf{V}}). \quad (46)$$

The proposed inexact algorithm consists of the following main steps. Fig. 4 gives a graphical illustration of the proposed algorithm.

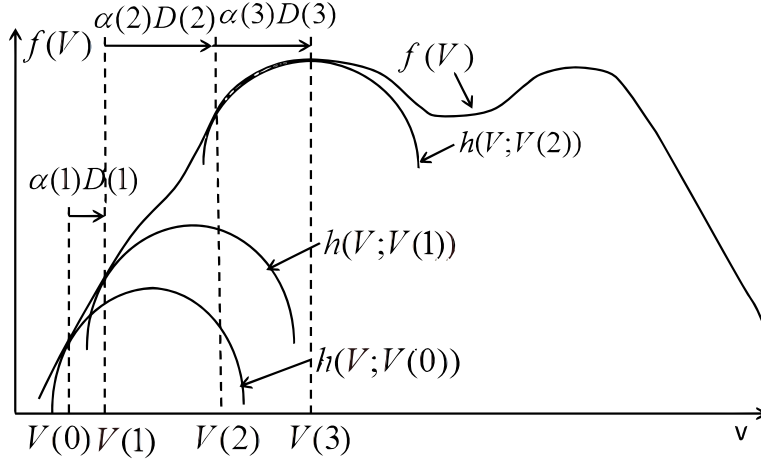


Fig. 4. A graphical illustration of the proposed Inexact-SCA algorithm, assuming $s(\mathbf{V})$ is not present. At $\mathbf{V}(0)$, a concave function $h(\mathbf{V}; \mathbf{V}(0))$ is used to approximate the original non-convex function $f(\mathbf{V})$. A good step that increases the objective value is taken to update the solution $\mathbf{V}(1)$. Then the concave function $h(\mathbf{V}; \mathbf{V}(1))$ is constructed at the point $\mathbf{V}(1)$, and so on. Continuing this process, a stationary solution of the original non-convex problem can be found.

Step 1 (Find update direction): Suppose $\mathbf{V}(t-1)$ is a feasible solution to (SYSTEM). At iteration t , we solve the following convex optimization problem for each cell $k \in \mathcal{K}$

$$\begin{aligned} \max_{\mathbf{D}^k} \quad & g_{\mathbf{D}^k}^k(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) + \frac{1}{2} \text{Tr}[(\mathbf{D}^k)^H \mathbf{G}^k(\mathbf{V}(t-1)) \mathbf{D}^k] - s^k(\mathbf{D}^k + \mathbf{V}^k(t-1)) \\ \text{s.t.} \quad & \mathbf{D}^k + \mathbf{V}^k(t-1) \in \mathcal{V}^k, \\ & \mathbf{D}^{q_k} + \mathbf{V}^{q_k}(t-1) \in \mathcal{V}^{q_k}, \quad \forall q_k \in \mathcal{Q}_k \end{aligned} \quad (\text{Q})$$

where $-\mathbf{G}^k(\mathbf{V}(t-1)) \succ 0$; $\text{Tr}[(\mathbf{D}^k)^H \mathbf{G}^k(\mathbf{V}(t-1)) \mathbf{D}^k]$ is an approximation of the second order directional derivative of $g^k(\cdot; \mathbf{V}(t-1))$ at point $\mathbf{V}(t-1)$. Let $\mathbf{D}^k(t)$ denote the optimal solution of the subproblem (Q).

Step 2 (Armijo step-size selection): For each cell k , choose the constants $\sigma \in (0, 1)$, $\alpha^{\text{init}} > 0$, $\beta^j \in (0, 1)$. Let $\alpha^k(t)$ be the largest element in $\{\alpha^{\text{init}}\beta^j\}_{j=0,1,\dots}$ satisfying:

$$u^k(\mathbf{V}^k(t-1) + \alpha^k(t)\mathbf{D}^k(t); \mathbf{V}(t-1)) \geq u^k(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) + \sigma \alpha^k(t) \left(g_{\mathbf{D}^k(t)}^k{}'(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) - s^k(\mathbf{V}^k(t-1) + \mathbf{D}^k) + s^k(\mathbf{V}^k(t-1)) \right). \quad (48)$$

Step 3 (Update precoder): Let $\mathbf{V}^k(t) = \mathbf{V}^k(t-1) + \alpha^k(t)\mathbf{D}^k(t)$, $\forall k \in \mathcal{K}$.

Step 4 (Update lower bound): For each user $i_k \in \mathcal{I}$, compute (24a)–(24d). Compute the updated lower bound function $h(\mathbf{V}; \mathbf{V}(t))$ according to (21).

Step 5: Let $t = t + 1$, go to Step 1.

Compared with the exact version of the SCA algorithm, at each iteration, the subproblem (Q) needs to be solved instead of the subproblem (Lower-Bound). One may wonder why this may be an easier task, as the new subproblem also appears to be a quadratic problem with many constraints. In fact, the flexibility provided here is the freedom to choose the matrix $\mathbf{G}^k(\mathbf{V})$. We will see shortly that as long as this matrix is chosen to be negative definite, the convergence of the algorithm is always guaranteed. This allows us to choose the matrix in a way that can further decompose the problem (Q) into $|\mathcal{Q}_k|$ subproblems (one for each BS q_k), each of which is potentially easy to solve. Of course, how this can be done is highly problem dependent, hence this issue will be discussed later with the applications.

Before analyzing the convergence of the proposed algorithm, we present two technical lemmas, the proofs for which can be found in Appendix B–C. The first lemma bounds the improvement of $u^k(\cdot; \widehat{\mathbf{V}})$ before and after the precoder has been updated (hence the improvement of the lower bound $h(\cdot; \widehat{\mathbf{V}}) - s(\cdot)$, cf. (46)). The second lemma shows that if $\mathbf{V}(t) = \mathbf{V}(t-1)$, then a stationary solution of problem (SYSTEM) is reached.

Lemma 2 Suppose the second order directional derivative of the concave function $g^k(\mathbf{V}^k; \widehat{\mathbf{V}})$ is lower bounded, that is, for all feasible $\mathbf{V}^k, \widehat{\mathbf{V}}$ and all feasible direction \mathbf{D}^k , there exists a constant $B^k > 0$

$$g_{\mathbf{D}^k}^k{}''(\mathbf{V}^k; \widehat{\mathbf{V}}) \geq -B^k \text{Tr}[(\mathbf{D}^k)^H \mathbf{D}^k]. \quad (49)$$

Then the following inequality is true for any $\alpha > 0$

$$\begin{aligned} & u^k(\mathbf{V}^k + \alpha \mathbf{D}^k; \widehat{\mathbf{V}}) - u^k(\mathbf{V}^k; \widehat{\mathbf{V}}) \\ & \geq \alpha g_{\mathbf{D}^k}^k{}'(\mathbf{V}^k; \widehat{\mathbf{V}}) - \frac{\alpha^2 B^k}{2} \text{Tr}[\mathbf{D}^k (\mathbf{D}^k)^H] - \alpha \left(s^k(\mathbf{V}^k + \mathbf{D}^k) - s^k(\mathbf{V}^k) \right) \end{aligned} \quad (50)$$

Lemma 3 If for a given precoder $\mathbf{V}(t-1)$, the optimal solution for the subproblem (Q) is $\mathbf{D}^k(t) = \mathbf{0}$ for each $k \in \mathcal{K}$. Then $\mathbf{V}(t-1)$ is a stationary solution for the sum-utility maximization problem (SYSTEM).

We then proceed to analyze the convergence property of the proposed algorithm. We leave the proof details of the following result to Appendix D.

Theorem 2 Assume the second order directional derivative of $g^k(\mathbf{V}^k; \hat{\mathbf{V}})$ is lower bounded as in (49). Further assume that $-\mathbf{G}^k(\mathbf{V}) \succ 0$ for all feasible \mathbf{V} and for all k , and that $\mathbf{G}^k(\mathbf{V})$ is continuous in \mathbf{V} . Then the inexact-SCA algorithm converges to a stationary solution of the problem (SYSTEM).

Remark 7 In the literature, algorithms related to the Inexact-SCA include the block successive convex approximation (BSCA) algorithm recently proposed in [62] and the coordinate gradient descent (CGD) method proposed in [67] (also see [64]). The BSCA (resp. CGD) computes a stationary solution of a non-convex problem with smooth objective function $f(\cdot)$ (resp. smooth plus separable nonsmooth function $f(\cdot) + s(\cdot)$) by solving a sequences of convex approximated subproblems. Like the inexact-SCA algorithm, the update direction is computed by solving a strictly convex problem, while the step-size of the update is determined via Armijo rule.

The inexact-SCA and the above mentioned two methods and their convergence results differ in several important places. On the one hand, the BSCA and CGD seem to be more general in that the variables can be updated in a block-by-block fashion. However, the inexact-SCA finds a good direction to improve the *lower bound* $h(\cdot) - s(\cdot)$ of the original objective, while both the BSCA and CGD try to improve the *original* objective directly. In the present application, the approach adopted by the inexact-SCA is more favorable. Thanks to the separability of the lower bound $h(\cdot) - s(\cdot)$, the improvement of its functional value before and after the update can be checked easily by each cell k (via checking $g^k(\cdot) - s^k(\cdot)$, cf. (48)). The same cannot be done if we were to check the original objective function. Moreover, if we were to adopt either BSCA or the CGD algorithm, it is not immediately clear how the subproblem for computing the update direction can be formulated.

B. Application to Linear Precoder Design in HetNet

Now that we have seen the proposed inexact-SCA algorithm and its convergence property, we proceed to demonstrate how it can be effectively utilized in precoder design problems. Let us take the linear precoder design problem discussed in Section IV-C as an example.

First of all, let us compute the first and second order directional derivative for $g^k(\mathbf{V}^k; \hat{\mathbf{V}})$ at the point $\hat{\mathbf{V}}^k$. Using the definition of $\hat{\mathbf{S}}_{i_k}$ and $\hat{\mathbf{J}}^k$ given in (22) and (40), we have

$$\begin{aligned} g_{\mathbf{D}^k}^k{}'(\hat{\mathbf{V}}^k; \hat{\mathbf{V}}) &= \sum_{q_k \in \mathcal{Q}_k} \sum_{i_k \in \mathcal{I}_k} \left(\text{Tr} \left[(\hat{\mathbf{S}}_{i_k}[q_k] - \sum_{p_k \in \mathcal{Q}_k} \hat{\mathbf{J}}^k[q_k, p_k] \hat{\mathbf{V}}_{i_k}^{p_k}) (\mathbf{D}_{i_k}^{q_k})^H \right] \right. \\ &\quad \left. + \text{Tr} \left[\mathbf{D}_{i_k}^{q_k} (\hat{\mathbf{S}}_{i_k}[q_k] - \sum_{p_k \in \mathcal{Q}_k} \hat{\mathbf{J}}^k[q_k, p_k] \hat{\mathbf{V}}_{i_k}^{p_k})^H \right] \right) \\ g_{\mathbf{D}^k}^k{}''(\hat{\mathbf{V}}^k; \hat{\mathbf{V}}) &= -2\text{Tr}[\mathbf{D}^{q_k} \hat{\mathbf{J}}^k (\mathbf{D}^{q_k})^H]. \end{aligned}$$

Let us choose the matrix $\widehat{\mathbf{G}}^k$ as $-2 \times \text{bkdlg}(\widehat{\mathbf{J}}^k)$, that is

$$\widehat{\mathbf{G}}^k = \begin{bmatrix} -2\widehat{\mathbf{J}}^k[1, 1], & \cdots, & \mathbf{0} \\ \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \cdots & -2\widehat{\mathbf{J}}^k[Q_k, Q_k] \end{bmatrix}. \quad (51)$$

Obviously this matrix is continuous in $\widehat{\mathbf{V}}$, as the matrix $\widehat{\mathbf{J}}^k$, defined in (22), is continuous in $\widehat{\mathbf{V}}$.

Note that other choices of $\widehat{\mathbf{G}}^k$ are also possible (e.g., let $\widehat{\mathbf{G}}^k = -\mathbf{I}$). However in practice choosing $\widehat{\mathbf{G}}^k$ in the form of (51) can effectively accelerate the convergence of the overall algorithm, as it represents an approximated version of the second directional derivative of $g^k(\mathbf{V}^k; \widehat{\mathbf{V}})$.

As there is no penalty term present, the objective of the subproblem (Q) at a point $\widehat{\mathbf{V}}$ can be written as

$$\begin{aligned} & \sum_{q_k \in \mathcal{Q}_k} \left(\sum_{i_k \in \mathcal{I}_k} \left(\text{Tr} \left[(\widehat{\mathbf{S}}_{i_k}[q_k] - \sum_{p_k \in \mathcal{Q}_k} \widehat{\mathbf{J}}^k[q_k, p_k] \widehat{\mathbf{V}}_{i_k}^{p_k}) (\mathbf{D}_{i_k}^{q_k})^H \right] \right. \right. \\ & \quad \left. \left. + \text{Tr} \left[\mathbf{D}_{i_k}^{q_k} (\widehat{\mathbf{S}}_{i_k}[q_k] - \sum_{p_k \in \mathcal{Q}_k} \widehat{\mathbf{J}}^k[q_k, p_k] \widehat{\mathbf{V}}_{i_k}^{p_k})^H \right] \right) - \sum_{i_k \in \mathcal{I}_k} \text{Tr} [\mathbf{D}_{i_k}^{q_k} \widehat{\mathbf{J}}^k[q_k, q_k] (\mathbf{D}_{i_k}^{q_k})^H] \right). \end{aligned}$$

For notational simplicity, let us define

$$\widehat{\mathbf{M}}^{q_k} \triangleq \text{bkdlg} \left\{ \widehat{\mathbf{S}}_{i_k}[q_k] - \sum_{p_k \in \mathcal{Q}_k} \widehat{\mathbf{J}}^k[p_k, q_k] \widehat{\mathbf{V}}_{i_k}^{p_k} \right\}_{i_k \in \mathcal{K}} \in \mathbb{C}^{M|\mathcal{I}_k| \times d_{i_k}}, \quad (52)$$

$$\widetilde{\mathbf{G}}^{q_k} = \begin{bmatrix} -2\widehat{\mathbf{J}}^k[q_k, q_k], & \cdots, & \mathbf{0} \\ \mathbf{0} & \ddots & \vdots \\ \mathbf{0} & \cdots & -2\widehat{\mathbf{J}}^k[q_k, q_k] \end{bmatrix} \in \mathbb{C}^{M|\mathcal{I}_k| \times M|\mathcal{I}_k|}. \quad (53)$$

When the feasible sets are characterized by the per-BS power constraints: $\mathcal{V}^{q_k} = \{\mathbf{V}^{q_k} : \text{Tr}[\mathbf{V}^{q_k} (\mathbf{V}^{q_k})^H] \leq \bar{P}^{q_k}\}$, the subproblem (Q) can be again decomposed into $|\mathcal{Q}_k|$ subproblems, one for each BS q_k :

$$\begin{aligned} \max_{\mathbf{Y}^{q_k}} \quad & \text{Tr} \left[\widehat{\mathbf{M}}^{q_k} (\mathbf{Y}^{q_k} - \widehat{\mathbf{V}}^{q_k})^H + (\mathbf{Y}^{q_k} - \widehat{\mathbf{V}}^{q_k}) (\widehat{\mathbf{M}}^{q_k})^H + (\mathbf{Y}^{q_k} - \widehat{\mathbf{V}}^{q_k})^H \widetilde{\mathbf{G}}^{q_k} (\mathbf{Y}^{q_k} - \widehat{\mathbf{V}}^{q_k}) \right] \\ \text{s.t.} \quad & \text{Tr}[\mathbf{Y}^{q_k} (\mathbf{Y}^{q_k})^H] \leq \bar{P}^{q_k}. \end{aligned} \quad (\text{VIBC-I})$$

where we have made the transformation $\mathbf{Y}^{q_k} = \widehat{\mathbf{V}}^{q_k} + \mathbf{D}^{q_k}$. This problem is again a quadratic problem with a single ball constraint. Let $\lambda^{q_k} \geq 0$ denote the Lagrangian multiplier associated with the power constraint. Using the same argument leading to (33), we can show that the optimal solution for the subproblem (VIBC-I) is

$$(\mathbf{Y}^{q_k})^* = \left(-\widetilde{\mathbf{G}}^{q_k} + (\lambda^{q_k})^* \mathbf{I}_{M|\mathcal{I}_k|} \right)^{-1} \left(\widehat{\mathbf{M}}^{q_k} - \frac{1}{2} \widetilde{\mathbf{G}}^{q_k} \widehat{\mathbf{V}}^{q_k} \right) \quad (54)$$

or equivalently,

$$(\mathbf{Y}_{i_k}^{q_k})^* = \left(2\widehat{\mathbf{J}}^k[q_k, q_k] + (\lambda^{q_k})^* \mathbf{I}_M \right)^{-1} \left(\widehat{\mathbf{S}}_{i_k}[q_k] - \sum_{p_k \in \mathcal{Q}_k \setminus q_k} \widehat{\mathbf{J}}^k[p_k, q_k] \widehat{\mathbf{V}}_{i_k}^{p_k} \right), \quad \forall i_k \in \mathcal{I}_k. \quad (55)$$

It follows that the optimal direction for BS q_k to update its precoder is $\mathbf{D}^{q_k} = \mathbf{Y}^{q_k} - \widehat{\mathbf{V}}^{q_k}$. After each BSs finishes the computation of the direction matrices $\{\mathbf{D}^{q_k}\}_{q_k \in \mathcal{Q}_k}$, the macro-BS can collect them via the backhaul link, and compute the stepsize α^k using the Armijo rule. In this case, the Armijo rule (48) is simplified as selecting the largest α^k in $\{\alpha^{\text{init}}\beta^j\}_{j=0,1,\dots}$, such that

$$g^k(\widehat{\mathbf{V}}^k + \alpha^k \mathbf{D}^k; \widehat{\mathbf{V}}) - g^k(\widehat{\mathbf{V}}^k; \widehat{\mathbf{V}}) \geq \sigma \alpha^k \left(g_{\mathbf{D}^k}^k{}'(\widehat{\mathbf{V}}^k; \widehat{\mathbf{V}}) \right) \quad (56)$$

The proposed inexact-SCA algorithm for the precoder design in HetNet with intra-cell ComP and inter-cell CB is detailed in Table V. It is easily seen that the condition specified in the statement of Theorem 2 is satisfied, that is:

$$g_{\mathbf{D}^k}^k{}''(\mathbf{V}^k; \widehat{\mathbf{V}}) = -2\text{Tr}[(\mathbf{D}^k)^H \widehat{\mathbf{J}}^k \mathbf{D}^k] \geq -2\rho(\widehat{\mathbf{J}}^k)\text{Tr}[(\mathbf{D}^k)^H \mathbf{D}^k], \quad (57)$$

with $\rho(\widehat{\mathbf{J}}^k)$ being upper bounded for any feasible $\mathbf{V} \in \mathcal{V}$. Then the convergence of the algorithm given in Table V is a straightforward consequence of Theorem 2.

TABLE V
THE INEXACT-SCA ALGORITHM FOR SOLVING (SYSTEM) WITH INTRA-CELL COMP AND INTER-CELL CB

S1): **Initialization** Obtain a feasible solution $\mathbf{V}_{i_k}(0)$ for all $i_k \in \mathcal{I}$
S2): For each BS k , compute $\mathbf{C}_{i_k}(t)$, $\mathbf{E}_{i_k}(t)$, $\mathbf{U}_{i_k}(t)$ and $c_{i_k}(t)$, according to (24a)–(24d), for all $i_k \in \mathcal{I}_k$
S3): For each BS k , compute the following
S3a): $\mathbf{J}^k(t) = \sum_{j_l \in \mathcal{I}} c_{j_l}(t) (\mathbf{H}_{j_l}^k)^H \mathbf{U}_{j_l}(t) (\mathbf{E}_{i_k}(t))^{-1} (\mathbf{U}_{j_l}(t))^H \mathbf{H}_{j_l}^k$
S3b): $\mathbf{S}_{i_k}(t) = c_{i_k}(t) (\mathbf{H}_{i_k}^k)^H \mathbf{U}_{i_k}(t) (\mathbf{E}_{i_k}(t))^{-1}$, $\forall i_k \in \mathcal{I}_k$.
S4): For each cell k , compute the precoders $\mathbf{V}^k(t)$ by the following steps:
 Compute $\mathbf{Y}^{q_k}(t)$ using (54), $\forall q_k \in \mathcal{Q}_k$;
 Let $\mathbf{D}^{q_k}(t) = \mathbf{Y}^{q_k}(t) - \mathbf{V}^{q_k}(t-1)$, $\forall q_k \in \mathcal{Q}_k$;
 Perform the Armijo line search (57) to determine $\alpha^k(t)$,
 Let $\mathbf{V}^{q_k}(t+1) = \mathbf{V}^{q_k}(t) + \alpha^k(t) \mathbf{D}^{q_k}(t)$, $\forall q_k \in \mathcal{Q}_k$
S5) Until some stopping criterion is met.

VI. NUMERICAL RESULTS

In this section we conduct numerical experiments to validate the effectiveness of the proposed algorithms. Both the exact and inexact versions of the SCA algorithm are tested for three main settings: 1) Multicell downlink linear precoder design (i.e., the IBC model); 2) HetNet downlink linear precoder design with inter-cell CB and intra-cell JP; 3) HetNet downlink joint clustering and linear precoder design with intra-cell partial ComP.

The general setup for the experiments are given as follows. We consider a multicell network of up to 10 cells. The distance of the centers of two adjacent cells is set to be 500 meters (representing a HetNet with densely deployed cells. See Fig. 5 for an illustration). Both the BSs and the users are randomly placed in each cell. Let $y_{i_k}^{q_\ell}$ denote the distance between BS q_ℓ and user i_k . The channel coefficients between user i_k and BS q_ℓ are modeled as zero mean circularly symmetric complex Gaussian vector with $(200/y_{i_k}^{q_\ell})^3 L_{i_k}^{q_\ell}$ as variance for both real and imaginary

dimensions, where $10 \log 10(L_{i_k}^{q_\epsilon}) \sim \mathcal{N}(0, 64)$ is a real Gaussian random variable modeling the shadowing effect. We set the noise power $\sigma_{i_k}^2 = 1$ for all i_k , set the power budget $P_{q_k} = P$ for all q_k . We define the total transmission power for cell k as $P_k^{\text{tot}} = P|\mathcal{Q}_k|$.

The stopping criteria are chosen as follows. The single time-scale algorithm as well as the outer loop of the double time-scale algorithm stop when $\frac{|u(t+1)-u(t)|}{|u(t)|} \leq 10^{-3}$. The inner loop of the two-time scale algorithm stops while the relative increase of the objective value for the related subproblem (i.e., problem (VIBC-BLK)) is less than 10^{-3} after performing one round of update by all the BSs in the cell.

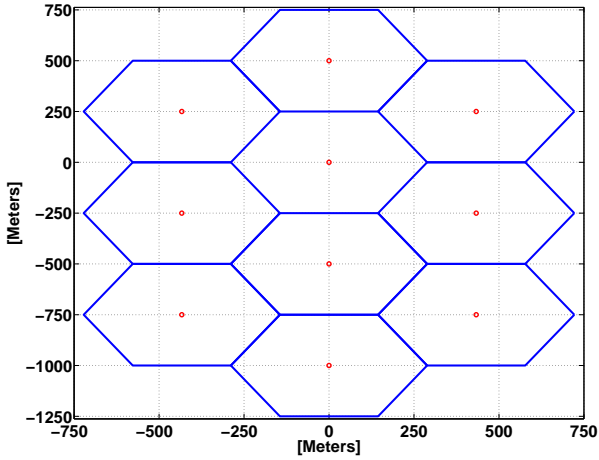


Fig. 5. Cell configuration for numerical experiments.

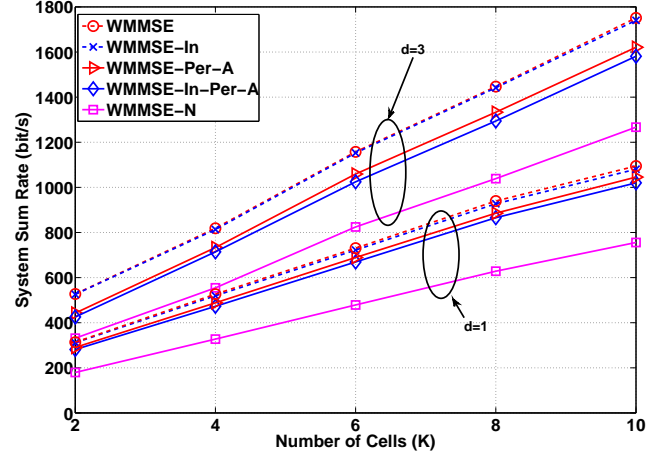


Fig. 6. Comparison of the system throughput of different algorithms in HetNet setting with different sizes of the network. $K = [2, 4, 6, 8, 10]$, $P_k^{\text{tot}} = 30\text{dB}$, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d_{i_k} = 1$ or $d_{i_k} = 3$ for all $i_k \in \mathcal{I}$.

A. HetNet and Multicell Downlink Setting

In this section, the performance of the following algorithms will be demonstrated and compared:

- 1) **WMMSE Algorithm** [44]: This algorithm is the one described in Table II. As this algorithm in its original form cannot deal with the per-BS power constraint in the HetNet setting, we will also consider its simple extension, in which the WMMSE algorithm is performed followed by a power normalization step for each BS $q_k \in \mathcal{Q}_k$. The latter algorithm is abbreviated as “WMMSE-N”;
- 2) **SCA-IN for IBC**: This algorithm is a simplification of that described in Table V for the IBC setting. It solves the same problem as the WMMSE algorithm;
- 3) **SCA for HetNet**: This algorithm is the two time-scale algorithm described in Table IV;
- 4) **SCA-IN for HetNet**: This algorithm is the inexact algorithm described in Table V;
- 5) **ZF-SCA for IBC**: This algorithm is the intra-cell ZF plus inter-cell CB algorithm described in Table III;
- 6) **Per-Cell ZF for HetNet** [21]: This algorithm is Algorithm 2 proposed in [21]. It performs the intra-cell ZF for the HetNet setting with BS power constraint, while completely ignoring the intra-cell interference.

All the algorithms considered in this subsection use the system sum rate as the users utility function. The plots to be shown represent the averaged performance of the algorithms running over 100 independent network generations.

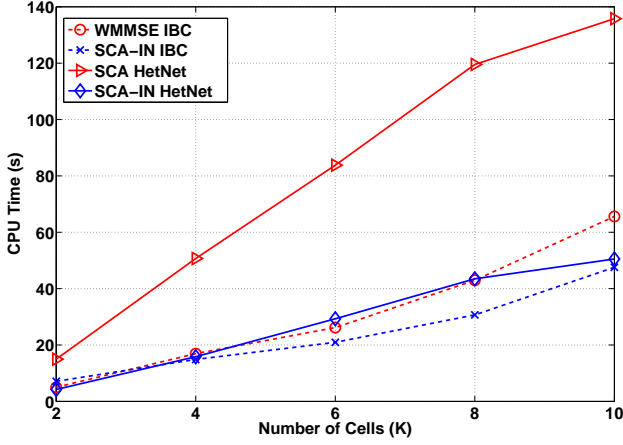


Fig. 7. Comparison of the CPU Time required for computation in HetNet setting with different sizes of the network. $K = [2, 4, 6, 8, 10]$, $P_k^{\text{tot}} = 30\text{dB}$, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d_{i_k} = 1$ for all $i_k \in \mathcal{I}$.

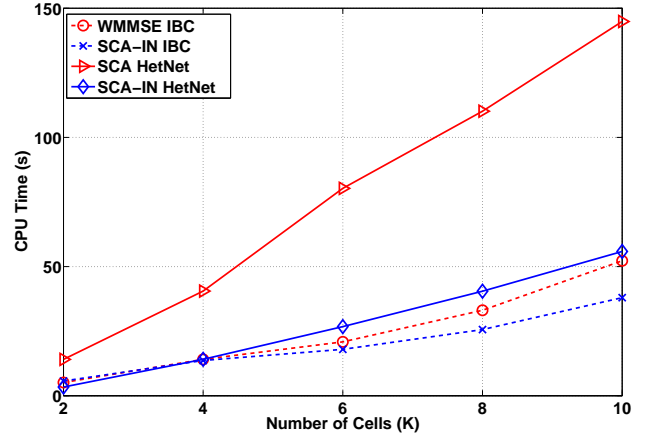


Fig. 8. Comparison of the CPU Time required for computation in HetNet setting with different sizes of the network. $K = [2, 4, 6, 8, 10]$, $P_k^{\text{tot}} = 30\text{dB}$, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d_{i_k} = 3$ for all $i_k \in \mathcal{I}$.

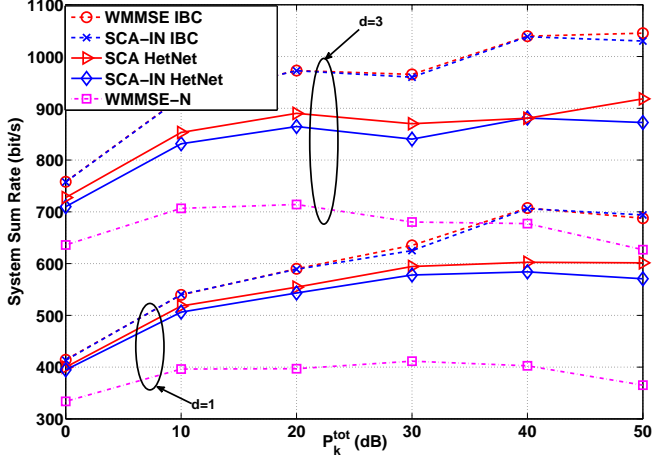


Fig. 9. Comparison of the system throughput of different algorithms in HetNet setting with different levels of transmit powers. $K = 5$, $P_k^{\text{tot}} = [0, 10, 20, 30, 40, 50]\text{dB}$, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d_{i_k} = 1$ or $d_{i_k} = 3$ for all $i_k \in \mathcal{I}$.

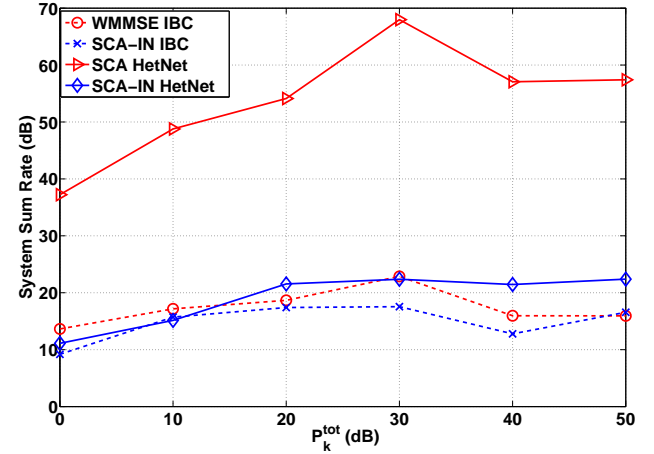


Fig. 10. Comparison of the CPU time needed for computation of different algorithms in HetNet setting with different levels of transmit powers. $K = 5$, $P_k^{\text{tot}} = [0, 10, 20, 30, 40, 50]\text{dB}$, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d_{i_k} = 1$ for all $i_k \in \mathcal{I}$.

Our first set of experiments compare the performance of the first four algorithms listed above. In Fig. 6–Fig. 8, the averaged system sum rate achieved by different algorithms as well as the averaged CPU time used is compared for a network with $|\mathcal{Q}_k| = 6$, $M = 5$, $N = 3$ and $|\mathcal{I}_k| = 10$. In Fig. 9–Fig. 11, a similar set of experiment is performed for networks with different SNR values. For the WMMSE and the “SCA-IN IBC” algorithm, the per-BS power constraint is completely ignored. Instead, a *single* per-cell power budget is assumed. Several interesting observations can be made. First of all the WMMSE algorithm and the SCA-IN algorithm in the IBC setting have almost identical performance in terms of both achieved sum rate and the computational time required. Secondly, in the HetNet setting the SCA-IN algorithm is much more efficient than the SCA algorithm. Remarkably, it uses the same computational resource as the WMMSE algorithm, while at the same time being able to enforce multiple per-BS power constraints. Additionally, the heuristic algorithm that combines the WMMSE with a simple normalization leads to poor sum rate performance.

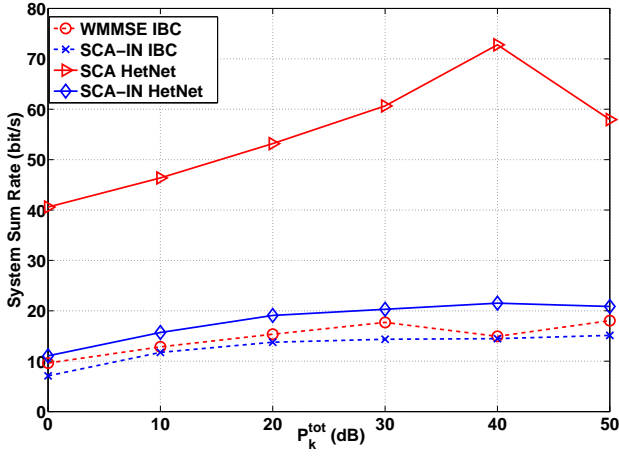


Fig. 11. Comparison of the CPU time needed for computation of different algorithms in HetNet setting with different levels of transmit powers. $K = 5$, $P_k^{\text{tot}} = [0, 10, 20, 30, 40, 50]$ dB, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = 10$, $M = 5$, $N = 3$, $d_{i_k} = 3$ for all $i_k \in \mathcal{I}$.

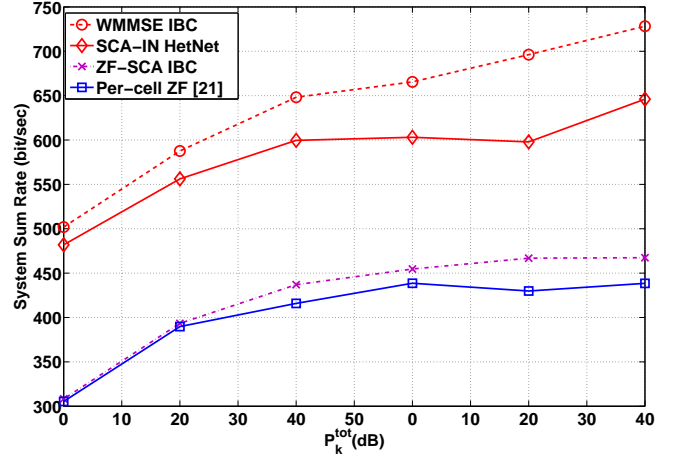


Fig. 12. Comparison of system sum rate achieved for different algorithms with different levels of transmit powers. $K = 5$, $P_k^{\text{tot}} = [0, 10, 20, 30, 40, 50]$ dB, $|\mathcal{Q}_k| = 5$, $|\mathcal{I}_k| = 10$, $M = 4$, $N = 2$, $d_{i_k} = 2$ for all $i_k \in \mathcal{I}$.

The second set of experiments compare the performance of the schemes that utilize general linear precoding and the ZF precoding. The results are summarized in Fig. 12–Fig. 14. In Fig. 12, we show the performance of the algorithms in the network with 5 cells and $|\mathcal{Q}_k| = 5$, $|\mathcal{I}_k| = 10$, $M = 4$, $N = 2$ and $d_{i_k} = 2$. Note that we have chosen the network parameters such that feasibility condition for ZF precoding is exactly satisfied, that is, $M|\mathcal{Q}_k| = N|\mathcal{I}_k|$. This is to say that for the ZF-based schemes, all the resources are dedicated to eliminating intra-cell interference, while the inter-cell interference is ignored. However, as suggested in Fig. 12, this is not an ideal strategy to deal with interference in densely deployed HetNet. The reason is that when the cells and the BSs are densely deployed, inter-cell interference is equally detrimental as the intra-cell interference on the users' rates. The general linear precoding approach, without pre-specifying which interference is to be cancelled, appears to be a more balanced way of dealing with the interference. This phenomenon is further highlighted in Fig. 13. In this figure, when $|\mathcal{I}_k|$ approaches the maximum number of allowable users for which the ZF strategy is still feasible (12 in this case, as $\frac{M|\mathcal{Q}_k|}{N} = 12$), the performance of both ZF based schemes drop sharply. On the other hand, when there is enough number of transmitters (or the BSs) in each cell, the ZF based-schemes perform as good as the general linear precoding based schemes.

B. Partial ComP in HetNet

In this section, we aim at jointly designing the clustering and linear precoding schemes in a partial ComP setting. To induce the desired clustering structure, we specialize the penalty terms in the objective to take the following form [54]: $s_{i_k}^{q_k}(\mathbf{V}) = \lambda \|\mathbf{V}_{i_k}^{q_k}\|_2$, $\forall i_k, q_k$ where $\lambda > 0$ is chosen appropriately to balance the resulting group size and the throughput performance. We compare the performance of the following three algorithms:

- 1) **WMMSE Algorithm:** This is our baseline algorithm that optimizes the precoders by treating all the BSs in each cell as a single virtual BS. The clustering structure is not optimized.

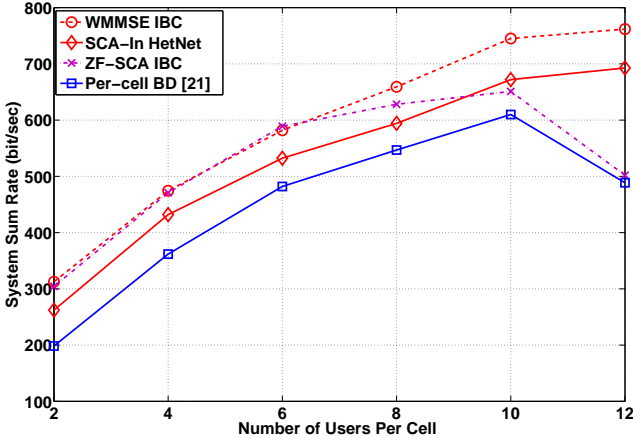


Fig. 13. Comparison of system sum rate achieved for different algorithms with different number of users per cell. $K = 5$, $P_k^{\text{tot}} = 30\text{dB}$, $|\mathcal{Q}_k| = 6$, $|\mathcal{I}_k| = [2, 4, 5, \dots, 12]$, $M = 4$, $N = 2$, $d_{i_k} = 2$ for all $i_k \in \mathcal{I}$.

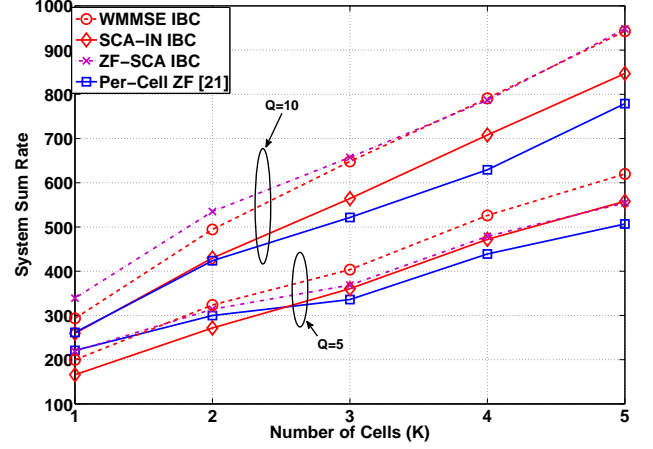


Fig. 14. Comparison of system sum rate achieved for different algorithms with different number of cells. $K = [1, 2, \dots, 5]$, $P_k^{\text{tot}} = 30\text{dB}$, $|\mathcal{Q}_k| = [5, 10]$, $|\mathcal{I}_k| = 10$, $M = 4$, $N = 2$, $d_{i_k} = 2$ for all $i_k \in \mathcal{I}$.

- 2) **SCA Algorithm:** This is the algorithm proposed in [54], which can be viewed as a special case of the SCA algorithm for solving the penalized utility maximization problem, see Remark 6.
- 3) **SCA-IN Algorithm:** This algorithm uses the inexact-SCA approach to solve the penalized utility maximization problem.

Our result is summarized in Fig. 15– Fig. 16 as well as in Table VI. We see that both the SCA and the SCA-IN based approach is able to keep a large portion of the system sum rate achieved by the full per-cell cooperation, while using only small cluster sizes. In contrast, the precoders generated by the WMMSE algorithm are void of any kind of clustering structure: they always mandate all the BSs to transmit to each user. Comparing the SCA and SCA-IN algorithm, we see that their performance is almost identical in terms of system throughput and the generated cluster sizes. However the advantage of using the inexact version is that it is computationally much more efficient than its exact counterpart.

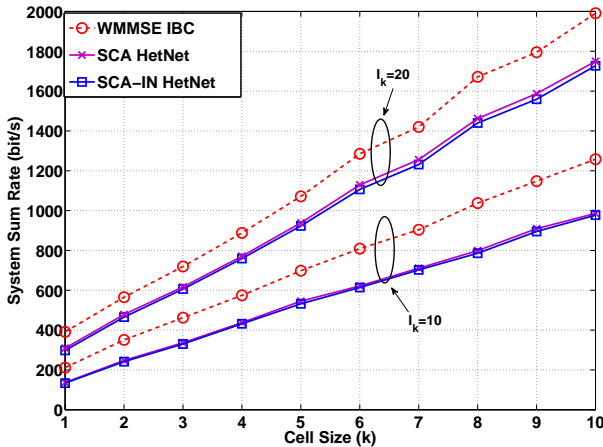


Fig. 15. Comparison of the system throughput of different algorithms in HetNet setting with different number of cells. $K = [1, 2, \dots, 10]$, $P_k^{\text{tot}} = 30\text{dB}$, $|\mathcal{Q}_k| = 10$, $|\mathcal{I}_k| = [10, 20]$, $M = 4$, $N = 2$, $d_{i_k} = 1$ for all $i_k \in \mathcal{I}$. $\lambda = 0.1$ when $|\mathcal{I}_k| = 10$, and $\lambda = 0.05$ when $|\mathcal{I}_k| = 20$.

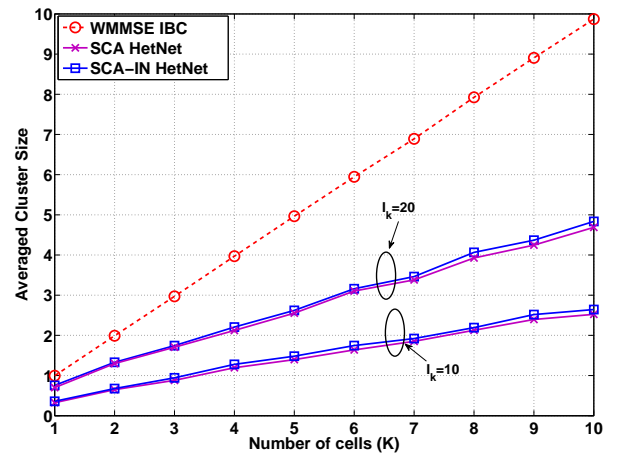


Fig. 16. Comparison of the averaged cluster size generated by different algorithms in HetNet setting with different number of cells. $K = [1, 2, \dots, 10]$, $P_k^{\text{tot}} = 30\text{dB}$, $|\mathcal{Q}_k| = 10$, $|\mathcal{I}_k| = [10, 20]$, $M = 4$, $N = 2$, $d_{i_k} = 1$ for all $i_k \in \mathcal{I}$. $\lambda = 0.1$ when $|\mathcal{I}_k| = 10$, and $\lambda = 0.05$ when $|\mathcal{I}_k| = 20$.

TABLE VI
CPU TIME NEEDED FOR DIFFERENT ALGORITHMS (UNIT: SECOND)

	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10
WMMSE ($I_k = 10$)	0.8	2.5	3.1	5.2	6.6	8.2	10.1	11.2	14.4	17.0
SCA ($I_k = 10$)	6.0	9.9	13.2	17.1	19.8	24.7	28.2	33.7	39.8	42.5
SCA-IN ($I_k = 10$)	0.8	1.7	2.7	3.4	4.4	6.0	6.9	8.5	9.8	11.8
WMMSE ($I_k = 20$)	2.6	6.5	11.0	15.6	21.7	28.1	36.0	44.5	52.4	62.5
SCA ($I_k = 20$)	20.6	28.3	34.4	45.5	57.1	70.8	80.9	87.4	106.7	121.3
SCA-IN ($I_k = 20$)	2.6	4.6	6.1	8.9	11.5	14.4	18.4	21.2	25.6	30.8

VII. CONCLUSION

In this paper we have addressed an important family of interference management problems arising in the heterogenous networks. The main novelty of this work lies in the proposal to achieve decomposition across the interference-coupled networks by using the technique of successive convex approximation. Our proposed approach is of low computational complexity, as each of the subproblems to be solved is convex and completely decomposes across the cells. Depending on the way that the subproblems are solved, two general algorithms have been proposed, both of which can be applied to many practical interference management problems. We believe that the framework studied in this paper is extendable to many other important problems for resource/interference management beyond those mentioned in this work.

VIII. ACKNOWLEDGEMENT

The first authors wishes thank Dr. Zi Xu from Shanghai University for helpful discussions in many technical issues in this paper.

APPENDIX

A. An Alternative Proof of Lemma 1

Proof: Note that we have $\mathbf{C}_{i_k} \succ 0$. Thus in order to show that $l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k})$ is jointly convex with respect to $(\mathbf{V}_{i_k}, \mathbf{C}_{i_k})$ in their respective domains, it is sufficient to show that for all $\mathbf{D}_{i_k} \in \mathbb{C}^{MQ_k \times d_{i_k}}$, $\mathbf{M}_{i_k} \in \mathbb{S}^N$ and t such that $\mathbf{C}_{i_k} + t\mathbf{M}_{i_k} \succ 0$, the second order directional derivative of $l_{i_k}(\mathbf{V}_{i_k}, \mathbf{C}_{i_k})$ is non-negative [60]

$$\frac{d^2 l_{i_k}(\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}, \mathbf{C}_{i_k} + t\mathbf{M}_{i_k})}{dt^2} \geq 0. \quad (58)$$

Let us first investigate the first order directional derivative. We have that

$$\begin{aligned}
& \frac{dl_{i_k}(\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}, \mathbf{C}_{i_k} + t\mathbf{M}_{i_k})}{dt} \\
&= \frac{d\text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}) \right]}{dt} \\
&= \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}) \right] \\
&\quad + \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k \mathbf{D}_{i_k} \right] \\
&\quad - \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}) \right] \\
&\triangleq a(t) + b(t) + c(t)
\end{aligned} \tag{59}$$

For each of the above three terms, we take a further order of derivative with respect to t . For the first term, we have:

$$\begin{aligned}
\frac{da(t)}{dt} &= \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k \mathbf{D}_{i_k} \right] \\
&\quad - \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}) \right]
\end{aligned} \tag{60}$$

Similarly, for the second term we have

$$\begin{aligned}
\frac{db(t)}{dt} &= \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k \mathbf{D}_{i_k} \right] \\
&\quad - \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k \mathbf{D}_{i_k} \right]
\end{aligned} \tag{61}$$

For the third term, we have

$$\begin{aligned}
\frac{dc(t)}{dt} &= \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}) \right] \\
&\quad + \text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k \mathbf{D}_{i_k} \right] \\
&\quad - 2\text{Tr} \left[(\widehat{\mathbf{E}}_{i_k})^{-1} (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k})^H (\mathbf{H}_{i_k}^k)^H (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}) \right]
\end{aligned} \tag{62}$$

For notational simplicity, let us define

$$\begin{aligned}
\mathbf{A}_{i_k} &\triangleq (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1/2} \mathbf{H}_{i_k}^k \mathbf{D}_{i_k} \\
\mathbf{B}_{i_k} &\triangleq (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1/2} \mathbf{M}_{i_k} (\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1} \mathbf{H}_{i_k}^k (\mathbf{V}_{i_k} + t\mathbf{D}_{i_k})
\end{aligned} \tag{63}$$

Note that we can write $(\mathbf{C}_{i_k} + t\mathbf{M}_{i_k})^{-1/2}$ because of the assumption, $(\mathbf{C}_{i_k} + t\mathbf{M}_{i_k}) \succ 0$.

Utilizing (60)–(62), we have the following expression for the second order derivative of $l_{i_k}(\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}, \mathbf{C}_{i_k} +$

$t\mathbf{M}_{i_k})$

$$\begin{aligned}
& \frac{d^2 l_{i_k}(\mathbf{V}_{i_k} + t\mathbf{D}_{i_k}, \mathbf{C}_{i_k} + t\mathbf{M}_{i_k})}{dt^2} \\
&= \frac{da(t)}{dt} + \frac{db(t)}{dt} + \frac{dc(t)}{dt} \\
&= \text{Tr} [\mathbf{E}_{i_k}^{-1} (2\mathbf{A}_{i_k}^H \mathbf{A}_{i_k} - 2\mathbf{A}_{i_k}^H \mathbf{B}_{i_k} - 2\mathbf{B}_{i_k}^H \mathbf{A}_{i_k} + 2\mathbf{B}_{i_k}^H \mathbf{B}_{i_k})] \\
&= 2\text{Tr} [\mathbf{E}_{i_k}^{-1} (\mathbf{A}_{i_k} - \mathbf{B}_{i_k})^H (\mathbf{A}_{i_k} - \mathbf{B}_{i_k})] \geq 0
\end{aligned} \tag{64}$$

where the last inequality is from the fact that $(\mathbf{A}_{i_k} - \mathbf{B}_{i_k})^H (\mathbf{A}_{i_k} - \mathbf{B}_{i_k}) \succeq 0$, $\mathbf{E}_{i_k} \succ 0$, and the fact that the trace of the product of two positive semi-definite matrices is nonnegative. This completes the proof. \blacksquare

B. Proof of Lemma 2

Proof: We have the following series of inequalities

$$\begin{aligned}
& u^k(\mathbf{V}^k + \alpha\mathbf{D}^k; \hat{\mathbf{V}}) - u^k(\mathbf{V}^k; \hat{\mathbf{V}}) \\
& \stackrel{(i)}{\geq} \alpha g_{\mathbf{D}^k}^k(\mathbf{V}^k; \hat{\mathbf{V}}) - \frac{\alpha^2 B^k}{2} \text{Tr}[\mathbf{D}^k (\mathbf{D}^k)^H] - \left(s^k(\mathbf{V}^k + \alpha\mathbf{D}^k) - s^k(\mathbf{V}^k) \right) \\
& \stackrel{(ii)}{\geq} \alpha g_{\mathbf{D}^k}^k(\mathbf{V}^k; \hat{\mathbf{V}}) - \frac{\alpha^2 B^k}{2} \text{Tr}[\mathbf{D}^k (\mathbf{D}^k)^H] - \left(\alpha s^k(\mathbf{V}^k + \mathbf{D}^k) + (1 - \alpha) s^k(\mathbf{V}^k) - s^k(\mathbf{V}^k) \right) \\
& = \alpha g_{\mathbf{D}^k}^k(\mathbf{V}^k; \hat{\mathbf{V}}) - \frac{\alpha^2 B^k}{2} \text{Tr}[\mathbf{D}^k (\mathbf{D}^k)^H] - \alpha \left(s^k(\mathbf{V}^k + \mathbf{D}^k) - s^k(\mathbf{V}^k) \right)
\end{aligned}$$

where (i) is from the well known Descent Lemma [68, Proposition A.32], (ii) is from the convexity of $s^k(\cdot)$, which is implied by the convexity of $s_{i_k}^{q_k}(\cdot)$ assumed in assumption B-1. \blacksquare

C. Proof of Lemma 3

Proof: The fact that the optimal solution problem (Q) is $\mathbf{D}^k(t) = \mathbf{0}$ implies that, for all feasible \mathbf{D}^k , we have

$$0 \geq g_{\mathbf{D}^k}^k(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) + \frac{1}{2} \text{Tr}[(\mathbf{D}^k)^H \mathbf{G}^k(\mathbf{V}(t-1)) \mathbf{D}^k] - s^k(\mathbf{V}^k(t-1) + \mathbf{D}^k) + s^k(\mathbf{V}^k(t-1)). \tag{65}$$

Clearly for any $\epsilon \in (0, 1)$, $\mathbf{D}^k = \epsilon\mathbf{D}^k + (1 - \epsilon)\mathbf{0}$ is also a feasible direction. Then for all $\epsilon \in (0, 1)$, we have

$$0 \geq g_{\epsilon\mathbf{D}^k}^k(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) + \frac{\epsilon^2}{2} \text{Tr}[(\mathbf{D}^k)^H \mathbf{G}^k(\mathbf{V}(t-1)) \mathbf{D}^k] - s^k(\mathbf{V}^k(t-1) + \epsilon\mathbf{D}^k) + s^k(\mathbf{V}^k(t-1)).$$

Using the concavity of the function $g^k(\cdot; \mathbf{V}(t-1))$, we obtain

$$\begin{aligned}
0 & \geq g^k(\mathbf{V}^k(t-1) + \epsilon\mathbf{D}^k; \mathbf{V}(t-1)) - g^k(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) \\
& \quad + \frac{\epsilon^2}{2} \text{Tr}[(\mathbf{D}^k)^H \mathbf{G}^k(\mathbf{V}(t-1)) \mathbf{D}^k] - s^k(\mathbf{V}^k(t-1) + \epsilon\mathbf{D}^k) + s^k(\mathbf{V}^k(t-1)).
\end{aligned}$$

Dividing both sides of the above inequality by $\epsilon > 0$, and letting ϵ goes to zero, we obtain, for all feasible \mathbf{D}^k ,

$$g_{\mathbf{D}^k}^{k'}(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) - s_{\mathbf{D}^k}^k(\mathbf{V}^k(t-1)) \leq 0. \quad (66)$$

An immediate consequence of this result is that for all feasible directions

$$h'_{\mathbf{D}}(\mathbf{V}(t-1); \mathbf{V}(t-1)) - s'_{\mathbf{D}}(\mathbf{V}(t-1)) = \sum_{k \in \mathcal{K}} g_{\mathbf{D}^k}^{k'}(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) - s_{\mathbf{D}^k}^k(\mathbf{V}^k(t-1)) \leq 0. \quad (67)$$

Utilizing (27), we obtain that for all feasible directions \mathbf{D} , $f'_{\mathbf{D}}(\mathbf{V}(t-1)) - s'_{\mathbf{D}}(\mathbf{V}(t-1)) \leq 0$, which says $\mathbf{V}(t-1)$ is a stationary solution of problem (SYSTEM). ■

D. Proof of Theorem 2

Proof: We first show that $\mathbf{D}^k(t)$ is in essence an “ascent direction” of the function $u^k(\mathbf{V}^k; \mathbf{V}(t-1))$, that is, we have

$$g_{\mathbf{D}^k(t)}^{k'}(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) - s^k(\mathbf{V}^k(t-1) + \mathbf{D}^k(t)) + s^k(\mathbf{V}^k(t-1)) > 0, \forall \mathbf{D}^k(t) \neq \mathbf{0}. \quad (68)$$

Clearly, $\mathbf{D}^k = \mathbf{0}$ is a feasible solution for problem (Q), which yields an objective value of $-s^k(\mathbf{V}^k(t-1))$. From the optimality of the solution $\mathbf{D}^k(t)$, it follows that

$$\begin{aligned} & g_{\mathbf{D}^k(t)}^{k'}(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) + \frac{1}{2} \text{Tr}[(\mathbf{D}^k(t))^H \mathbf{G}^k(\mathbf{V}(t-1)) \mathbf{D}^k(t)] \\ & - s^k(\mathbf{V}^k(t-1) + \mathbf{D}^k(t)) + s^k(\mathbf{V}^k(t-1)) \geq 0. \end{aligned}$$

This inequality combined with the strict negative definiteness of the matrix $\mathbf{G}^k(\mathbf{V}(t-1))$ further implies that

$$\begin{aligned} & g_{\mathbf{D}^k(t)}^{k'}(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) - s^k(\mathbf{V}^k(t-1) + \mathbf{D}^k(t)) + s^k(\mathbf{V}^k(t-1)) \\ & \geq -\frac{1}{2} \text{Tr}[(\mathbf{D}^k(t))^H \mathbf{G}^k(\mathbf{V}(t-1)) \mathbf{D}^k(t)] \geq \frac{b^k}{2} \text{Tr}[(\mathbf{D}^k(t))^H \mathbf{D}^k(t)] > 0, \forall \mathbf{D}^k(t) \neq \mathbf{0}. \end{aligned} \quad (69)$$

where $b^k > 0$ is a constant such that $-\mathbf{G}^k(\mathbf{V}(t-1)) \succeq b_k \mathbf{I}$.

This shows (68). When the Armijo rule (48) is used for step-size selection, the inequality (68) implies that

$$\begin{aligned} & u^k(\mathbf{V}^k(t); \mathbf{V}(t-1)) - u^k(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) \\ & \geq \sigma \alpha^k(t) \left(g_{\mathbf{D}^k(t)}^{k'}(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) - s^k(\mathbf{V}^k(t-1) + \mathbf{D}^k(t)) + s^k(\mathbf{V}^k(t-1)) \right) \geq 0. \end{aligned} \quad (70)$$

Using the lower bound property (20a)–(20b), we obtain (cf. the derivation in (25))

$$\begin{aligned} u(\mathbf{V}(t)) &= f(\mathbf{V}(t)) - s(\mathbf{V}(t)) \geq \sum_{k \in \mathcal{K}} u^k(\mathbf{V}^k(t); \mathbf{V}(t-1)) \\ &\geq \sum_{k \in \mathcal{K}} u^k(\mathbf{V}^k(t-1); \mathbf{V}(t-1)) = u(\mathbf{V}(t-1)). \end{aligned} \quad (71)$$

The above series of inequalities show that $\{u(\mathbf{V}(t))\}$ is a monotonically increasing and convergent sequence. This fact combined with (70) implies that

$$\lim_{t \rightarrow \infty} \sigma \alpha^k(t) \left(g_{\mathbf{D}^k(t)}^k \left(\mathbf{V}^k(t-1); \mathbf{V}(t-1) \right) - s^k \left(\mathbf{V}^k(t-1) + \mathbf{D}^k(t) \right) + s^k \left(\mathbf{V}^k(t-1) \right) \right) = 0. \quad (72)$$

We then claim that in the limit, the direction $\mathbf{D}^k(t)$ converges to $\mathbf{0}$. Assume the contrary, then there must exist a constant $\delta > 0$ such that $\liminf_{t \rightarrow \infty} \|\mathbf{D}^k(t)\| = \delta$. Then there must exist an infinite subsequence $\{t_r\}_{r=1}^\infty$ such that $\lim_{r \rightarrow \infty} \|\mathbf{D}^k(t_r)\| = \delta > 0$. Then we show that along such subsequence, the stepsize $\alpha^k(t_r)$ is lower bounded, that is

$$\exists c^k > 0 \text{ s.t. } 0 < c^k \leq \alpha^k(t_r) \leq 1, \forall r. \quad (73)$$

Let us suppose that at iteration t_r , the ℓ -th trial of the line search is successful. Then according to the Armijo rule, in the $(\ell-1)$ -th trial, we must have

$$\begin{aligned} u^k(\mathbf{V}^k(t_r-1) + \alpha^{\text{init}} \beta^{\ell-1} \mathbf{D}^k(t_r); \mathbf{V}(t_r-1)) &< u^k(\mathbf{V}^k(t_r-1); \mathbf{V}(t_r-1)) \\ &+ \sigma \alpha^{\text{init}} \beta^{\ell-1} \left(g_{\mathbf{D}^k(t_r)}^k \left(\mathbf{V}^k(t_r-1); \mathbf{V}(t_r-1) \right) - s^k \left(\mathbf{V}^k(t_r-1) + \mathbf{D}^k(t_r) \right) + s^k \left(\mathbf{V}^k(t_r-1) \right) \right) \end{aligned}$$

which is equivalent to

$$\begin{aligned} u^k(\mathbf{V}^k(t_r-1) + \alpha^k(t_r) \beta^{-1} \mathbf{D}^k(t_r); \mathbf{V}(t_r-1)) &< u^k(\mathbf{V}^k(t_r-1); \mathbf{V}(t_r-1)) \\ &+ \sigma \alpha^k(t_r) \beta^{-1} \left(g_{\mathbf{D}^k(t_r)}^k \left(\mathbf{V}^k(t_r-1); \mathbf{V}(t_r-1) \right) - s^k \left(\mathbf{V}^k(t_r-1) + \mathbf{D}^k(t_r) \right) + s^k \left(\mathbf{V}^k(t_r-1) \right) \right). \end{aligned}$$

From Lemma 2, we can further obtain

$$\begin{aligned} \alpha^k(t_r) g_{\mathbf{D}^k(t_r)}^k \left(\mathbf{V}^k(t_r-1); \mathbf{V}(t_r-1) \right) &- \frac{(\alpha^k(t_r))^2 B^k}{2} \text{Tr}[\mathbf{D}^k(t_r)(\mathbf{D}^k(t_r))^H] - \alpha^k(t_r) (s^k(\mathbf{V}^k(t_r-1) + \mathbf{D}^k(t_r)) - s^k(\mathbf{V}^k(t_r-1))) \\ &< \sigma \alpha^k(t_r) \beta^{-1} \left(g_{\mathbf{D}^k(t_r)}^k \left(\mathbf{V}^k(t_r-1); \mathbf{V}(t_r-1) \right) - s^k(\mathbf{V}^k(t_r-1) + \mathbf{D}^k(t_r)) + s^k(\mathbf{V}^k(t_r-1)) \right). \end{aligned}$$

Rearranging terms, and utilizing (69), we obtain

$$\begin{aligned} \alpha^k(t_r) &\geq \frac{2\beta(1-\sigma) \left(g_{\mathbf{D}^k(t_r)}^k \left(\mathbf{V}^k(t_r-1); \mathbf{V}(t_r-1) \right) - s^k \left(\mathbf{V}^k(t_r-1) + \mathbf{D}^k(t_r) \right) + s^k \left(\mathbf{V}^k(t_r-1) \right) \right)}{B^k \text{Tr}[(\mathbf{D}^k(t_r))^H \mathbf{D}^k(t_r)]} \\ &> \frac{2\beta(1-\sigma)b^k}{B^k} > 0 \end{aligned} \quad (74)$$

where the last inequality is from the fact that $B^k > 0$, $b^k > 0$, and $\sigma \in (0, 1)$.

Taking a further subsequence of $\{t_r\}$ if necessary, let $\{\mathbf{V}(t_r)\}$ be a converging subsequence, with limit point \mathbf{V}^* . From the Armijo step-size selection rule and the series of inequalities in (71), we must have

$$\begin{aligned}
& u(\mathbf{V}(t_r)) - u(\mathbf{V}(t_r - 1)) \\
& \geq \sum_{k \in \mathcal{K}} \left(u^k(\mathbf{V}^k(t_r); \mathbf{V}(t_r - 1)) - u^k(\mathbf{V}^k(t_r - 1); \mathbf{V}(t_r - 1)) \right) \\
& \geq \sum_{k \in \mathcal{K}} \sigma \alpha^k(t_r) \left(g_{\mathbf{D}^k(t_r)}^k \left(\mathbf{V}^k(t_r - 1); \mathbf{V}(t_r - 1) \right) - s^k \left(\mathbf{V}^k(t_r - 1) + \mathbf{D}^k(t_r) \right) + s^k \left(\mathbf{V}^k(t_r - 1) \right) \right) \\
& \geq \sum_{k \in \mathcal{K}} \sigma \alpha^k(t_r) \frac{b^k}{2} \text{Tr} \left[(\mathbf{D}^k(t_r))^H \mathbf{D}^k(t_r) \right].
\end{aligned}$$

The fact that $\{u(\mathbf{V}(t))\}$ converges implies that

$$\lim_{r \rightarrow \infty} \sum_{k \in \mathcal{K}} \sigma \alpha^k(t_r) b^k \text{Tr} \left[(\mathbf{D}^k(t_r))^H \mathbf{D}^k(t_r) \right] = 0.$$

This is only possible when $\lim_{r \rightarrow \infty} \text{Tr} \left[(\mathbf{D}^k(t_r))^H \mathbf{D}^k(t_r) \right] = 0$, which contradicts to the assumption of $\lim_{r \rightarrow \infty} \mathbf{D}^k(t_r) > \mathbf{0}$. As a result, we conclude that $\lim_{t \rightarrow \infty} \mathbf{D}^k(t) = \mathbf{0}$.

Next we claim that every subsequence of $\{\mathbf{V}(t)\}$ converges to a stationary solution of problem (SYSTEM). Let $\{\mathbf{V}(t_\ell)\}$ be a converging subsequence, with limit point \mathbf{V}^* . As $\mathbf{D}^k(t_\ell + 1)$ is the solution for the convex problem (Q), then for all feasible \mathbf{D}^k

$$\begin{aligned}
& g_{\mathbf{D}^k(t_\ell+1)}^k \left(\mathbf{V}^k(t_\ell); \mathbf{V}(t_\ell) \right) + \frac{1}{2} \text{Tr} \left[(\mathbf{D}^k(t_\ell + 1))^H \mathbf{G}^k(\mathbf{V}(t_\ell)) \mathbf{D}^k(t_\ell + 1) \right] - s^k \left(\mathbf{V}^k(t_\ell) + \mathbf{D}^k(t_\ell + 1) \right) \\
& \geq g_{\mathbf{D}^k}^k \left(\mathbf{V}^k(t_\ell); \mathbf{V}(t_\ell) \right) + \frac{1}{2} \text{Tr} \left[(\mathbf{D}^k)^H \mathbf{G}^k(\mathbf{V}(t_\ell)) \mathbf{D}^k \right] - s^k \left(\mathbf{V}^k(t_\ell) + \mathbf{D}^k \right).
\end{aligned} \tag{75}$$

Taking limit on both sides, by the assumed continuity of the directional derivative of $g^k(\cdot)$ and the continuity of $\mathbf{G}^k(\mathbf{V})$, as well as the fact $\lim_{t \rightarrow \infty} \mathbf{D}^k(t) = \mathbf{0}$, we obtain that for all feasible \mathbf{D}^k

$$0 \geq g_{\mathbf{D}^k}^k \left((\mathbf{V}^k)^*; \mathbf{V}^* \right) + \frac{1}{2} \text{Tr} \left[(\mathbf{D}^k)^H \mathbf{G}^k(\mathbf{V}^*) \mathbf{D}^k \right] - s^k \left((\mathbf{V}^k)^* + \mathbf{D}^k \right) + s^k \left((\mathbf{V}^k)^* \right). \tag{76}$$

Following the same proof as in Lemma 3, we can show that for all feasible directions \mathbf{D} , $f_{\mathbf{D}}'(\mathbf{V}^*) - s_{\mathbf{D}}'(\mathbf{V}^*) \leq 0$, which says \mathbf{V}^* is a stationary solution of problem (SYSTEM). ■

REFERENCES

- [1] 3GPP, “Evolved Universal Terrestrial Radio Access (EUTRA) and Evolved Universal Terrestrial Radio Access Network (EUTRAN); overall description,” 2011, 3GPP TS 36.300, V8.9.0.
- [2] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, “A survey on 3GPP heterogeneous networks,” *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, 2011.
- [3] V. Chandrasekhar and J.G. Andrews, “Femtocell networks: A survey,” *IEEE Communications Magazine*, pp. 59–67, 2008.
- [4] M. Hong and Z.-Q. Luo, “Signal processing and optimal resource allocation for the interference channel,” *EURASIP E-Reference Signal Processing*, 2012, accepted, available at <http://arxiv.org>.

- [5] A. S. Avestimehr, H. El Gamal; S. A. Jafar, S. Ulukus, and S. Vishwanath Ed., “IEEE Transactions on Information Theory,” 2011, Special Issue on Interference Networks.
- [6] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1380–1408, 2010.
- [7] G.J. Foschini, K. Karakayali, and R.A. Valenzuela, “Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency,” *IEE Proceedings Communications*, vol. 153, no. 4, pp. 548–555, 2006.
- [8] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Communications Magazine*, , no. 2, pp. 102–111, 2011.
- [9] S.-J. Kim and G.B. Giannakis, “Optimal resource allocation for MIMO Ad Hoc Cognitive Radio Networks,” *IEEE Transactions on Information Theory*, vol. 57, no. 5, pp. 3117–3131, 2011.
- [10] C. Shi, D. A. Schmidt, R. A. Berry, M. L. Honig, and W. Utschick, “Distributed interference pricing for the MIMO interference channel,” in *IEEE International Conference on Communications, 2009*, june 2009, pp. 1–5.
- [11] Z. K. M. Ho and D. Gesbert, “Balancing egoism and altruism on interference channel: The MIMO case,” in *2010 IEEE International Conference on Communications (ICC)*, may 2010, pp. 1–5.
- [12] E. Larsson and E. Jorswieck, “Competition versus cooperation on the MISO interference channel,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, pp. 1059–1069, 2008.
- [13] E. Jorswieck and E. Larsson, “The MISO interference channel from a game-theoretic perspective: A combination of selfishness and altruism achieves pareto optimality,” in *IEEE ICASSP*, april 2008, pp. 5364–5367.
- [14] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, “Coordinated beamforming for MISO interference channel: Complexity analysis and efficient algorithms,” *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1142–1157, 2011.
- [15] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna Gaussian broadcast channel,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.
- [16] W. Yu and T. Lan, “Transmitter optimization for the multi-antenna downlink with per-antenna power constraints,” *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2646–2660, 2007.
- [17] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, “Sum power iterative water-filling for multi-antenna Gaussian broadcast channels,” *IEEE Transactions on information theory*, vol. 51, no. 4, pp. 1570–1580, 2005.
- [18] W. Yu, “Sum-capacity computation for the gaussian vector broadcast channel via dual decomposition,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 754–759, 2006.
- [19] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, 2004.
- [20] J. Zhang, R. Chen, J.G. Andrews, A. Ghosh, and R.W. Heath, “Networked MIMO with clustered linear precoding,” *IEEE Transactions on Wireless Communications*, , no. 8, pp. 1910–1921, 2009.
- [21] R. Zhang, “Cooperative multi-cell block diagonalization with per-base-station power constraints,” *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1435–1445, 2010.
- [22] S. S. Christensen, R. Agarwal, E. D. Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, 2008.
- [23] Z.-Q. Luo and S. Zhang, “Dynamic spectrum management: Complexity and duality,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 57–73, 2008.
- [24] M. Razaviyayn, M. Hong, and Z.-Q. Luo, “Linear transceiver design for a MIMO interfering broadcast channel achieving max-min fairness,” 2012, Submitted to Signal Processing.
- [25] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo, “Linear transceiver design for interference alignment: Complexity and computation,” *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2896–2910, 2012.
- [26] Z.-Q. Luo and W. Yu, “An introduction to convex optimization for communications and signal processing,” *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1426–1438, 2006.

- [27] D.P. Palomar and Mung Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1439–1451, 2006.
- [28] S. Ye and R.S. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2839–2848, 2003.
- [29] C. Shi, R. A. Berry, and M. L. Honig, "Monotonic convergence of distributed interference pricing in wireless networks," in *Proceedings of the 2009 IEEE international conference on Symposium on Information Theory - Volume 3*, 2009, ISIT'09, pp. 1619–1623.
- [30] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 4020–4030, 2008.
- [31] P. Tsiaflakis, M. Diehl, and M. Moonen, "Distributed spectrum management algorithms for multiuser DSL networks," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4825–4843, 2008.
- [32] P. Tsiaflakis, I. Necoara, J. Suykens, and M. Moonen, "Improved dual decomposition based optimization for DSL dynamic spectrum management," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2230–2245, 2010.
- [33] M. Hong and Z.-Q. Luo, "Joint linear precoder optimization and base station selection for an uplink MIMO network: A game theoretic approach," in *the Proceedings of the IEEE ICASSP*, 2012.
- [34] M. Hong, J. Garzás, A. García-Armada, and A. Garcia, "Lower bounds optimization for coordinated linear transmission beamformer design in multicell network downlink," Submitted to *IEEE Transactions on Signal Processing*.
- [35] J. Papandriopoulos and J. S. Evans, "SCALE: A low-complexity distributed protocol for spectrum balancing in multiuser DSL networks," *IEEE Transactions on Information Theory*, vol. 55, no. 8, 2009.
- [36] F. Wang, M. Krunz, and S. G. Cui, "Price-based spectrum management in cognitive radio networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, 2008.
- [37] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Transactions on Communications*, vol. 54, no. 7, pp. 1310–1322, 2006.
- [38] L. Venturino, N. Prasad, and X. Wang, "Coordinated scheduling and power allocation in downlink multicell OFDMA networks," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 6, pp. 2835–2848, 2009.
- [39] L. Venturino, N. Prasad, and X. Wang, "Coordinated linear beamforming in downlink multicell wireless networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1451–1461, 2010.
- [40] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 145–152, 2004.
- [41] Z.-Q. Luo and J.-S. Pang, "Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–10, 2006.
- [42] G. Scutari, D. P. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: A unified view," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 7, 2008.
- [43] G. Scutari, D. P. Palomar, and S. Barbarossa, "Asynchronous iterative water-filling for Gaussian frequency-selective interference channels," *IEEE Transactions on Information Theory*, vol. 54, no. 7, 2008.
- [44] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, 2011.
- [45] S. Verdú, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.
- [46] T. M. Cover and J. A. Thomas, *Elements of Information Theory, second edition*, Wiley, 2005.
- [47] H. Boche and M. Schubert, "A calculus for log-convex interference functions," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5469–5490, 2008.
- [48] S. Stanczak, M. Wiczanowski, and H. Boche, "Distributed utility-based power control: Objectives and algorithms," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 5058–5068, 2007.
- [49] H. Boche, S. Naik, and T. Alpcan, "Characterization of convex and concave resource allocation problems in interference coupled wireless systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2382–2394, 2011.

- [50] H. Boche and M. Schubert, "A unifying approach to interference modeling for wireless networks," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3282–3297, 2010.
- [51] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Transactions Wireless Communications*, vol. 6, no. 7, pp. 2640–2651, 2007.
- [52] W. Yu, "Multiuser water-filling in the presence of crosstalk," in *Information Theory and Applications Workshop 07*, 2007.
- [53] C. T. K. Ng and H. Huang, "Linear precoding in cooperative MIMO cellular networks with limited coordination clusters," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1446–1454, 2010.
- [54] M. Hong, R. Sun, and Z.-Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogenous networks," 2012, accepted by IEEE Journal on Selected Areas in Communications, Special issue on Large Scale Multi-antenna Systems.
- [55] S.-H. Park, O. Simeone, O. Sahin, and S. Shamai, "Robust and efficient distributed compression for cloud radio access networks," 2012, available online at Arxiv.org.
- [56] S.-J. Kim, S. Jainand, and G.B. Giannakis, "Backhaul-constrained multi-cell cooperation using compressive sensing and spectral clustering," in *2012 IEEE SPWAC*, 2012.
- [57] Y.-F. Liu, Y.-H. Dai, and Z.-Q. Luo, "Max-min fairness linear transceiver design for a multi-user MIMO interference channel," in *the Proceedings of the international conference on Communicaitons 2011*, 2011.
- [58] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "Linear transceiver design for a MIMO interfering broadcast channel achieving max-min fairness," in *2011 Asilomar Conference on Signals, Systems, and Computers*, 2011.
- [59] M. R. Garey and D. S. Johnson, *Computers and Intractability: A guide to the Theory of NP-completeness*, W. H. Freeman and Company, San Francisco, U.S.A, 1979.
- [60] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [61] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.21," Apr. 2011.
- [62] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," 2012, Submitted to *SIAM Journal on Optimization*.
- [63] David G. Luenberger, *Linear and Nonlinear Programming, Second Edition*, Springer, 1984.
- [64] P. Tseng and S. Yun, "Block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization," *Journal of Optimization Theory and Applications*, vol. 140, pp. 513–535, 2009.
- [65] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *Journal of Optimization Theory and Applications*, vol. 103, no. 9, pp. 475–494, 2001.
- [66] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.
- [67] Paul Tseng and Sangwoon Yun, "A coordinate gradient descent method for nonsmooth separable minimization," *Mathematical Programming*, vol. 117, pp. 387–423, 2009, 10.1007/s10107-007-0170-0.
- [68] D. P. Bertsekas, *Nonlinear Programming, 2nd ed*, Athena Scientific, Belmont, MA, 1999.